The role of the PWWP domain in the DNA methyltransferase 3A targeting to the genome

Gintarė Sendžikaitė

Magdalene College

Babraham Institute

University of Cambridge

February 2021

This thesis is submitted for the degree of

Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Data, figures, and research published as scientific articles by the author during doctoral studies have been included in the thesis.

In accordance with the Faculty of Biology guidelines, this thesis does not exceed 60,000 words.

Signed:

Date: 25/02/2021

Gintarė Sendžikaitė

Cambridge

Summary

All cells in a mammalian body share the same DNA content but have a distinct appearance and function. This is achieved by activation and repression of genes by epigenetic mechanisms, which are crucial for stable differentiation of cells in a multi-cellular organism. DNA methylation, established by DNA methyltransferase (DNMT) enzymes, regulates gene expression and is essential for mammalian development. Histone tail post-translational modifications modulate the recruitment and activity of DNMTs. In particular, the PWWP domains of DNMT3A and DNMT3B are posited to interact with histone 3 lysine 36 trimethylation (H3K36me3). In this study, I characterise a mouse model carrying a D329A point mutation in the DNMT3A PWWP domain, predicted to ablate the DNMT3A-PWWP and H3K36me3 interaction. Similar mutations in humans cause primordial dwarfism. I found that oocytes of *Dnmt3a*^{D329A} mice show no alterations in DNA methylation. However, the D329A mutation causes dominant postnatal growth retardation in mice. At the molecular level, it results in aberrant progressive acquisition of DNA methylation across domains marked by H3K27me3 and bivalent chromatin in multiple adult tissues. These domains characteristically contain important developmental regulatory genes, and their aberrant methylation is associated with de-repression of these genes in the adult hypothalamus. Recapitulation of this mutation in mouse embryonic stem cells (mESCs) did not identify new DNMT3A^{D329A} interacting partners. Upon differentiation of *Dnmt3a*^{D329A} carrying mESCs to neuronal progenitor cells, a minor gain in DNA methylation is observed. This work provides key molecular insights into the role of DNMT3A in regulating postnatal growth, the targeting of DNMT3A to the genome, function of the DNMT3A PWWP domain and its association with histone modifications.

Acknowledgements

First, I would like to thank my supervisor, Gavin Kelsey, for giving me an opportunity to pursue this work and supporting me along the way. I am forever grateful for your excellent scientific advice and unlimited patience.

Thanks to all the members of Kelsey lab who shared their knowledge and insights with me. In particular, thanks to Elena Ivanova for generating preliminary data and giving me a head start with the project. A very special thanks goes to my mentor and dear friend Courtney Hanna for offering extensive advice on matters of work and life at any time of the day.

Thanks to the amazing people at Babraham Institute with whom I shared the office, the lab, and many memorable moments. This work would not have been possible without Biological Support Unit, Sequencing facility, Bioinformatics team and other technical staff at Babraham Institute. Thanks to MRC DTP programme for funding my project and my secondments.

I am very grateful to Tuncay Baubec who hosted me in his lab at University of Zurich. Thanks to members of Baubec lab, in particular Massimiliano Manzo, Joel Wirz, Sara Giulliani, for sharing their expertise in the lab and for showing me how to float down the Limmat on an inflatable island.

Huge thanks to my friends for showing endless support despite not quite knowing what it is I study. I am especially grateful to Steve, for encouraging me to pursue a PhD in the first place, and for sharing infinite scientific discussions at home.

Galiausiai, ačiū mano šeimai už meilę, nenuilstantį rūpestį ir moralinį palaikymą mano akademiniame kelyje. Ačiū, mama, už tavo neblėstantį tikėjimą manimi.

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

ADD	ATRX-DNMT3-DNMT3L
BAH	bromo-adjacent homology domain
BirA	biotin ligase
CENP-C	centromere protein C
CGI	CpG island
ChIP	chromatin immunoprecipitation
CpH	cytosine followed by either A, C or T nucleotide
DEG	differentially expressed gene
DMAP1	DNA methyltransferase associated protein 1
DMAPD	DMAP1-interacting domain
DMR	differentially methylated region
DNMT	DNA methyltransferase
E	embryonic day
EB	elution buffer
ESC	embryonic stem cell
GH	growth hormone
GO	gene ontology
GV	germinal vesicle
H3K4me3	histone 3 lysine 4 trimetyhlation
HDAC	histone deacetylase

HP1	heterochromatin protein 1
HSV-Tk	Herpes Simplex Virus-1 thymidine kinase ganciclovir-sensitivity
Hy	hygromycin-resistance
ICR	imprinting control region
Igf	Insulin-like growth factor
lncRNA	Long non-coding RNA
MBD	methyl-binding domain
MeCP2	methyl-CpG-binding protein 2
NLS	nuclear localisation signal
NPC	neuronal progenitor cells
P	Postnatal day
PBAT	post-bisulphite adaptor tagging
PCA	principal component analysis
PGC	primordial germ cells
PRC2	polycomb repressive complex 2
PTM	post-translational modification
PWWP	Pro-Trp-Trp-Pro motif
RAMBiO	recombinase-assisted mapping of biotin-tagged proteins
RFTS	replication foci targeting sequence
RMCE	recombinase-mediated exchange
RPKM	Reads per kilobase per million mapped reads

RT	room temperature
SAM	S-adenosyl L-methionine
TBRS	Tatton-Brown-Rahman syndrome
TET	ten-eleven translocation
TSS	transcription start site
UHRF1	Ubiquitin-like, PHD and ring finger-containing 1
Δ	deletion

List of Appendices

1 Introduction

Declaration of author contributions

An earlier version of this chapter has been published:

Sendžikaitė, G. and Kelsey, G. (2019). The role and mechanisms of DNA methylation in the oocyte. Essays in Biochemistry, 63 (6): 691–705.
https://doi.org/10.1042/EBC20190043

1.1 Epigenetic regulation of the genome

A mammalian body is made of cells and tissues that have a wide range of appearances and functions, although they all contain exactly the same genetic material. Cell-specific properties are acquired by activation and/or repression of a variety of genes. Epigenetic mechanisms, defined as changes to genome function or chromosome properties that are heritable but do not alter the underlying DNA code, regulate gene expression by altering chromatin accessibility to transcription factors, which allows cell-specific gene activation (Bernstein et al., 2006). There are two major epigenetic mechanisms controlling the genome – DNA methylation and histone tail modification, and the interplay between the two enable complexity in the regulatory system to emerge. Generally, epigenetic regulation of gene expression allows lineage specification during development and is required for stable differentiation in multicellular organisms. These changes can be heritable at a cellular level during proliferation and, in some cases, can be passed on through generations, therefore facilitating an intergenerational inheritance of non-genetic cues (Xavier et al., 2019).

1.2 Histone code

In all eukaryotic cells DNA is compactly packaged into the cell nucleus by formation of chromatin. Approximately 146bp of DNA are wrapped around a nucleosome, the basic unit of chromatin. The nucleosome is a scaffold protein complex, containing two of each H2A, H2B, H3 and H4 histones in an octamer (Richmond et al., 1984). Protruding C- and N-terminal tails of histones can acquire post-translational modifications (PTMs), which can alter how compactly multiple nucleosomes might come together. Modifications of histone tails are associated with distinct functions, such as opening or closing of the chromatin, and recruitment of interactor proteins. For example, actively transcribed genes have histone 3 lysine 4 trimethylation (H3K4me3) enrichment at the promoter region and H3K36me3 enrichment over the transcribed gene body (Bannister et al., 2005; Santos-Rosa et al., 2002). In addition to PTMs, there are variants of key histones, which tend to be prevalent in particular contexts, such as after DNA damage and during non-replicative state changes (Henikoff and Smith, 2015). DNA methylation and histone modifications together control chromatin form and function.

1.3 DNA methylation

DNA is a double helix made of adenine – thymine, and cytosine – guanine nucleotide pairs held by Watson-Crick bonds (Watson and Crick, 1953). In addition to the four base nucleotides, there are three chemically modified nucleotides: C5-methylcytosine, N4-methylcytosine and N6-methyladenine (Ehrlich and Wang, 1981; Jeltsch, 2002). DNA methyltransferase enzymes (DNMTs) use a methyl- group from the donor S-adenosyl L-methionine (SAM) and covalently attach it to the nucleotide (Ehrlich and Wang, 1981). The methyl- group is positioned in the major DNA groove, therefore leaving Watson-Crick bond intact but affecting allosteric conformation of the double helix and binding of DNA-interacting proteins. The extent of global DNA methylation is variable in different mammalian tissues and at different developmental stages, and was suggested to play a role in development and tissue differentiation from the very early days of epigenetic research (Ehrlich et al., 1982; Gama-Sosa et al., 1983).

In prokaryotes, both cytosine and adenine methylation is found. However, unlike in mammals, these modifications are used in cell cycle-related gene expression, mismatch DNA repair and as a defence mechanism against bacteriophage invasion (Jeltsch, 2002). In mammalian DNA, only C5-methylcytosine is prevalent and is usually, but not

exclusively, found within a CpG dinucleotide context (Feng et al., 2010; Roy and Weissbach, 1975). This creates a symmetric pattern, where palindromic CG dyads are methylated on both DNA strands (Bird, 1978). After replication, the newly synthesised strand is unmethylated, but the hemimethylated palindrome can be recognised by the maintenance DNMT1, which methylates the nascent strand (Hermann et al., 2004). This way a symmetric pattern is passed to daughter cells during cell division allowing for a mechanism of inheritance (Bird, 2002).

Overall, DNA methylation is considered to be a repressive mark, especially at heterochromatin, pericentromeric regions, gene promoters, repetitive and transposable elements (Schübeler, 2015). In contrast, methylation over gene bodies is associated with active transcription. Functionally, DNA methylation alters binding of transcription factors and other chromatin interacting proteins, chromatin structure and accessibility, thus fine-tuning gene expression (Schübeler, 2015).

The major focus of this thesis is targeting of mammalian DNA methylation enzyme DNMT3A to the genome, thus, hereon the focus is on mammalian systems. As well as the main model of this thesis, the majority of the studies described below were conducted in murine systems or mammalian cell culture.

1.4 CpG distribution

Generally, CpG dinucleotides are underrepresented in vertebrate genomes, which show high levels of DNA methylation (Bird, 1980). One of the explanations for this is an increased chance of spontaneous deamination of methylated cytosine, creating a transition from cytosine to thymine (Ehrlich et al., 1986; Holliday and Grigg, 1993). Normally, deamination of unmethylated cytosine to uracil is rapidly repaired by uracil-DNA glycosylase, which cleaves newly incorporated uracil (Ehrlich et al., 1990). In contrast, methylcytosine mutates to thymine, and relies on mismatch repair machinery to remove the newly mutated nucleotide, with a risk that the wild-type guanine on a sister strand is cleaved by chance (Neddermann and Jiricny, 1993). Nonetheless, approximately 75% of CpG sites are methylated in mammalian genomes (Feng et al., 2010). A notable exception to this is CpG islands (CGIs); clusters showing high CpG dinucleotide density, high conservation and generally low levels of DNA methylation (Bird et al., 1985; Illingworth et al., 2010). Many CGIs found within gene promoters and transcription start

sites (TSS) are constitutively unmethylated, but some exhibit a lineage-specific DNA methylation status, shaping the transcriptional landscape.

1.5 DNA methylation machinery

The DNA methyltransferase enzyme family in mammals consists of five members: one maintenance, three *de novo* methyltransferases, and a co-factor (Figure 1.1).

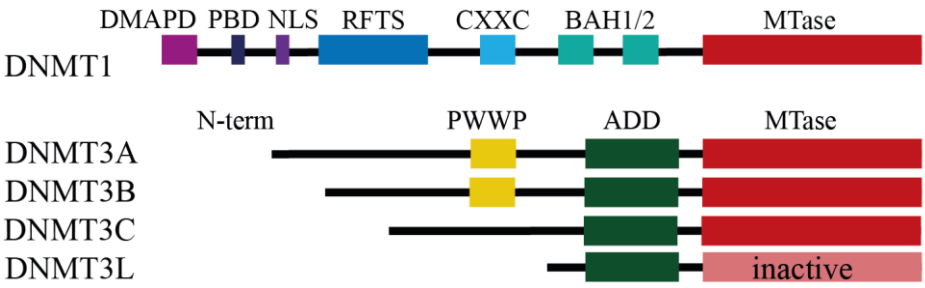


Figure 1.1. Schematic illustration of mammalian DNMT protein family. DMAPD – DMAP1-interacting domain, PBD – PCNA binding domain, NLS – nuclear localisation signal, RFTS – replication foci targeting sequence, BAH – bromo-adjacent homology domain, MTase – methyltransferase, PWWP – Pro-Trp-Trp-Pro motif, ADD – ATRX-DNMT3-DNMT3L domain.

DNMT1, the maintenance DNMT, recognises and methylates the unmethylated strand on hemimethylated DNA (Hermann et al., 2004). Homozygous deletion of *Dnmt1* results in embryonic lethality, and mice fail to develop further than embryonic day 8.5/9 (E8.5-E9) (Li et al., 1992). DNMT1 is a large protein containing a regulatory N-terminal domain and conserved methyltransferase domain. At the very N-terminus there is a DMAP1-interacting domain (DMAPD) required for interaction with DNA methyltransferase associated protein 1 (DMAP1) and Histone Deacetylase 2 (HDAC2) for transcriptional repression (Rountree et al., 2000). A second small domain is Proliferating Cell Nuclear Antigen (PCNA) binding domain, which leads to DNMT1 recruitment to the replication fork, followed by a nuclear localisation signal (NLS) (Chuang et al., 1997; Suetake et al., 2006). The Replication Foci Targeting Sequence (RFTS) domain is required for DNMT1 localisation to hemimethylated sites through Ubiquitin-like, PHD and Ring finger-containing 1 (UHRF1) protein at S phase (Bostick et al., 2007; Sharif et al., 2007). The

nearby CXXC domain is involved in an autoinhibitory methyltransferase domain control, together with bromo-adjacent homology domains (BAH1 and BAH2) (Song et al., 2012, 2011).

In addition to DNMT1, there are three *de novo* DNMTs, which use unmethylated DNA as a substrate. DNMT3s show high similarity and share some of the domains, further described in Section 1.11. DNMT3A and DNMT3B exhibit partial redundancy and are both required for setting up epigenetic states during embryogenesis (Dahlet et al., 2020; Okano et al., 1999). *Dnmt3a*$^{-/-}$ mice fail to survive longer than 3 weeks postnatally, and die for unknown reasons. Meanwhile *Dnmt3b*$^{-/-}$ is embryonic lethal and developmental defects arise at E10.5, characterised by neural tube formation defects and growth restriction (Okano et al., 1999). While there is some redundancy, generally *Dnmt3b* is considered to be more important during embryogenesis, as exemplified by a more severe phenotype in the knock-out (Borgel et al., 2010; Okano et al., 1999). Double *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$ mutant mice die before E11.5 but developmental defects are thought to arise immediately following gastrulation and can be observed at E8.5 (Dahlet et al., 2020; Okano et al., 1999). DNMT3C, a third, murine-specific *de novo* methyltransferase, has recently been discovered; it silences evolutionary young retrotransposons in prospermatogonia by methylating their promoters (Barau et al., 2016; Jain et al., 2017).

DNMT3L is the odd member of DNMT family, since it does not have an active catalytic domain (Aapola et al., 2000). *Dnmt3l*$^{-/-}$ mice are viable but sterile. It acts as an essential co-factor for establishment of genomic imprints in sperm and oocyte (Bourc'his et al., 2001). DNMT3L is also less conserved between the species (Yokomine et al., 2006). The C-terminal domain of DNMT3L can bind DNMT3A and DNMT3B C-terminal domains and significantly enhances their chromatin binding and/or catalytic activity (Chen et al., 2005; Hata et al., 2002; Jia et al., 2007). This might occur through formation of tetramers or conformational changes leading to increased affinity for methyl- donor SAM binding (Kareta et al., 2006; Gowher et al., 2005). This stimulatory effect appears to be ubiquitous for DNMT3A, but not for DNMT3B (Chédin et al., 2002; Veland et al., 2019). Meanwhile, DNMT3L does not appear to have an effect on DNMT1 enzymatic activity (Suetake et al., 2004).

1.6 Epigenetic changes during development

While in differentiated adult tissues DNA methylation is stably maintained, the ability of chromatin to undergo dynamic transitions is especially important during gamete and early embryo development, when two major epigenetic reprogramming waves are observed in mammals (Figure 1.2). Reprogramming is required to abolish established patterns determining cell lineage and to restore pluripotency (Cantone and Fisher, 2013; Lee et al., 2014; Seisenberger et al., 2012).

Both sperm and egg are terminally differentiated cells. After fertilisation, the embryonic cells of the new embryo undergo epigenetic reprogramming to erase the gamete-specific epigenome and regain totipotency (Xu and Xie, 2018). During pre-implantation development, parental DNA is rapidly demethylated. By the time the inner cell mass of the blastocyst is formed at E3.5, the DNA methylation and histone PTM patterns that characterised the gametes are almost completely lost and only a small subset of gamete differentially methylated regions (gDMRs) and histone PTMs are retained (Borgel et al., 2010; Smith et al., 2012; Xu and Xie, 2018). After implantation, from the epiblast stage to gastrula (in the mouse between embryonic days E4.5 to E6.5), DNA methylation is regained and established in a lineage-specific pattern (Borgel et al., 2010).

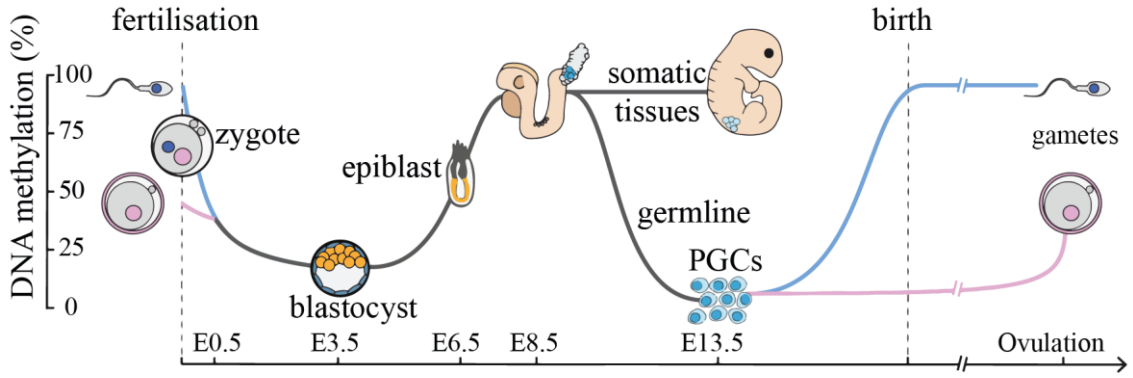


Figure 1.2. DNA methylation dynamics in development. After fertilisation, DNA methylation patterns specific to sperm (blue) and egg (pink) are erased and DNA methylation is re-established in epiblast at around embryonic day 6.5 (E6.5). Somatic tissues acquire further lineage and tissue specific methylation. Meanwhile, primordial germ cells (PGCs) undergo another epigenetic erasure and their DNA methylation is established in germ line just before birth is sperm and after reproductive maturation in females. Image reproduced under CC BY-SA 4.0 license from original by Marius Walter.

In mice, primordial germ cells (PGCs) are specified at E6.5-E7.25 in the yolk sac endoderm. As PGCs proliferate, they migrate to the genital ridge and undergo erasure of global DNA methylation (Guibert et al., 2012; Seisenberger et al., 2012). A second epigenetic reprogramming event is observed in these cells at E10.5-E11.5, when parent-of-origin epigenetic marks are erased, and PGCs differentiate into prospermatogonia or oogonia, depending on gonadal sex. Gamete-specific epigenomes are established in the germline soon after birth in male and in adulthood in female mice. This is accompanied by global transcriptional and epigenetic changes that are crucial for successful fertilisation and later development (Bonnet-Garnier et al., 2013).

1.7 DNA methylation in imprinting

DNA methylation appears to be dispensable for oocyte development and competence, as genetic ablation of oocyte methylation allows for successful fertilisation and ensuing embryonic development until the mid-gestation stage (Bourc'his et al., 2001; Branco et al., 2016; Hata et al., 2002; Kaneda et al., 2010, 2004). However, it is essential for genomic imprinting: a subset of the gamete specific differentially methylated regions (gDMRs) that evade embryonic reprogramming during early development result in parent-of-origin specific gene expression of the associated genes – imprinted genes. Failure to establish imprints leads to lethality or congenital diseases in both mice and humans, and the pathologies observed are linked to foetal growth, brain, and metabolic function. Interestingly, only three imprinted loci are conferred by methylation in male gametes, while there are at least 26 DNA methylation-dependent imprinted regions conferred in the oocyte (Kobayashi et al., 2012; Smallwood et al., 2011; Tucci et al., 2019). Imprinted loci contain a single or a cluster of genes, whose expression patterns are determined by imprinting control region (ICR) DNA methylation status. DNA methylation at ICRs is set up during oocyte growth, in a transcription-dependent manner. Disruption of transcription at the relevant loci leads to failure of DNA methylation establishment over ICRs and affects viability and monoallelic expression of imprinted gene loci in mice (Chotalia et al., 2009). Oocyte ICRs are composed of CGIs enriched for a specific sequence motif, recognised by KRAB zinc-finger protein 57 (ZFP57). ZFP57 protects imprinted sites from demethylation during embryonic reprogramming by recruiting KAP1, SETDB1, HP1 and NP95 to form a robustly silenced locus (Li et al., 2008; Quenneville et al., 2011). Imprinted genes exhibit stable and heritable monoallelic parent-of-origin specific gene expression that can be maintained throughout the

lifecourse, and is only overwritten during PGC reprogramming (Li and Li, 2019; Tucci et al., 2019).

Apart from ICRs, some gDMRs show transient or tissue-specific inheritance post fertilisation (Proudhon et al., 2012; Rutledge et al., 2014). Maternal non-imprinted gDMRs were shown to play a role in regulation of placental development in mice and a number of such gDMRs persist in human placenta and confer monoallelic expression of associated genes (Branco et al., 2016; Sanchez-Delgado et al., 2016).

1.8 DNA methylation in transcriptional regulation

Classically DNA methylation is considered a repressive, silencing modification particularly in gene promoter context and the relative density of CpG methylation at a promoter correlates with expression strength (Hsieh, 1994; Siegfried et al., 1999). The establishment of methyl-CpGs at promoter regions masks recognition sites of methylation-sensitive transcription factors and therefore prevent transcription (Yin et al., 2017). DNA methylation can also contribute to genomic stability by silencing of transposable elements, retroviruses, satellite repeats and telomeres (Barau et al., 2016; Gonzalo et al., 2006; Kuramochi-Miyagawa et al., 2008; Walsh et al., 1998; Xu et al., 1999).

In mammals, one of the X chromosomes of females has to be silenced to control the expression dosage of X chromosome genes between the sexes. The silencing of one copy of X chromosome is initiated by a *cis* acting long non-coding RNA (lncRNA) *Xist*, expressed from the X copy which is to be silenced (Clemson et al., 1996). *Xist* coats the X chromosome and recruits other chromatin silencing and condensation factors, such as polycomb repressor complex 2 (PRC2) which establishes a silencing histone mark H3K27me3 over the inactive chromosome (Zhao et al., 2008). The expression of *Xist* in somatic tissues is controlled by DNA methylation: the active X chromosome has a 5' DNA methylation of *Xist* which results in repression of this lncRNA, meanwhile the silenced X chromosome contains unmethylated *Xist* (Norris et al., 1994).

Methyl-binding domain (MBD) proteins are recruited by methylated cytosines and execute silencing via transcriptional repression domains (Barr et al., 2007; Hendrich and Bird, 1998; Stirzaker et al., 2017). MBDs can subsequently reinforce silencing by recruitment of other repressors or chromatin remodelling complexes, such as Sin3A and

NuRD (Boeke et al., 2000; Du et al., 2015; Jones et al., 1998; Nan et al., 1998; Ng et al., 2000; Zhang et al., 1999). Similarly, DNMT3 enzymes are able to recruit histone interacting proteins, such as HDAC1, to silence genomic regions independently of their DNA methylation function (Deplus et al., 2002; Fuks et al., 2001). DNMT3B has been shown to interact with HDAC1, HDAC2, Heterochromatin protein 1 (HP1), Suv39h1 and chromatin remodeller hSNF2H (Geiman et al., 2004). *Vice versa,* chromatin remodeller LSH can recruit DNMTs and HDACs to alter local chromatin environment and silence transcription at target loci (Clements et al., 2012; Han et al., 2020; Myant et al., 2011; Xi et al., 2009). Lysine methyltransferase G9A recruits DNA methylation to silence germline specific loci during embryogenesis (Auclair et al., 2016). Meanwhile, centromere protein C (CENP-C) and Suv39h-HP1 recruit DNMT3B to methylate peri- and centromeric regions (Gopalakrishnan et al., 2009; Lehnertz et al., 2003).

Notably, the methylation-transcription dynamic is reversed at gene bodies, where higher levels of DNA methylation correlate with higher gene expression (Lister et al., 2009). There are two main functions of intragenic DNA methylation: regulation of splicing and control of faithful transcription initiation. Methylation differentially marks main exons in relation to introns and changes in DNA methylation levels can impact alternative exon usage (Gelfman et al., 2013; Shayevitch et al., 2018). This could be an evolutionary mechanism of fast and heritable adaptation to the environment (Shukla et al., 2011). A second role of DNA methylation over gene bodies is to avert RNA polymerase II binding to orphan transcription start sites and to prevent aberrant transcription initiation and alternative promoter usage in different cell lineages (Maunakea et al., 2010; Neri et al., 2017).

1.9 Non-CpG DNA methylation

Non-CpG (CpH) methylation is detected in almost all human tissues (Schultz et al., 2015) but is present at a high level in oocytes, embryonic stem cells (ESCs) and some postmitotic tissues (Lister et al., 2009; Ramsahoye et al., 2000; Shirane et al., 2013; Tomizawa et al., 2011). In mouse ESCs, the levels of CpH methylation tend to decrease upon differentiation. In contrast, a marked increase in CpH methylation is observed in the brain during postnatal neuronal maturation (Lister et al., 2013). CpH methylation in the brain is restricted to vertebrates and is highly linked to the CpH reader methyl-CpG-binding protein 2 (MeCP2) (Guo et al., 2014; Mendoza et al., 2021). The dynamics and localisation of CpH methylation follow CpG methylation: neuro- and ESC- specific CpH

patterns can be attributed to DNMT3A and DNMT3B expression levels and localisation in these tissues (Lee et al., 2017; Ziller et al., 2011). CpH methylation cannot be maintained by DNMT1 in proliferating cells, and therefore is likely only functional in post-mitotic cells (Guo et al., 2014; Ramsahoye et al., 2000).

1.10 Demethylation

Demethylation of 5mC is of particular importance during development, when germ cell DNA methylation has to be erased in early embryogenesis, and during germ cell specification. There are two main ways for DNA demethylation: passive and active, and both of these are employed to alter the methylome of maternal and paternal genomes in development (Gu et al., 2011; F. Guo et al., 2014; Mayer et al., 2000; Oswald et al., 2000; Shen et al., 2014; Wossidlo et al., 2011). Passive DNA demethylation occurs when methylated sites are not maintained during cell replication and are therefore diluted (Kagiwada et al., 2013; Ohno et al., 2013; Rougier et al., 1998). A mechanistic example of passive dilution is exclusion of DNMT1 or its interacting partners UHRF1 or STELLA from the nucleus (Cardoso and Leonhardt, 1999; Du et al., 2019; Mulholland et al., 2020).

Alternatively, cytosine methylation can be catalysed by ten-eleven translocation (TET) methylcytosine dioxygenases, creating three intermediates in transition from 5mC to unmodified cytosine: 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Figure 1.3) (Ito et al., 2011). Any of these intermediates can be passively diluted during cell division or actively excised by Thymine DNA Glycosylase and subsequently replaced by unmethylated cytosine by the base mismatch repair machinery (Cortellino et al., 2011; Hajkova et al., 2010; He et al., 2011).

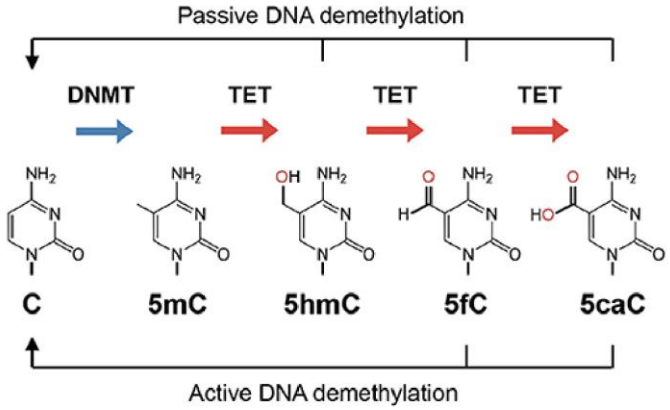


Figure 1.3. (De)methylation of cytosine. Intermediate forms of modified cytosine nucleotides and enzymes catalysing transition are shown. Reproduced from (Lio and Rao, 2019).

1.11 Recruitment of *de novo* DNMTs

Many studies have sought to understand how DNA methylation patterns are established in different genomic contexts. *De novo* methyltransferases have distinct intrinsic enzymatic sequence biases, where DNMT3A prefer TNC[G/A]CC and DNMT3B prefer TAC[G/A]GC, and are sensitive to CpG flanking nucleotides, but DNA sequence is not a sole determining factor of DNA methylation status (Jurkowska et al., 2011; Mallona et al., 2020; Mao et al., 2020; Wienholz et al., 2010). N-terminal regulatory domains of DNMT3s – the ADD (ATRX-DNMT3-DNMT3L) and PWWP (Pro-Trp-Trp-Pro motif) – can interact with various histone PTMs, which guide DNMT localisation and enzymatic activity.

1.11.1 The N-terminal domain and H3K27me3

The main divergence between DNMT3A and DNMT3B is a less studied intrinsically disordered N-terminal domain. The N-terminal domain of DNMT3A/B is responsible for non-specific DNA binding, association with nucleosomes, and contributes towards nuclear localisation (Jeong et al., 2009; Suetake et al., 2004). The DNMT3A N-terminal domain targets it to the shore of bivalent chromatin, harbouring developmental genes (Manzo et al., 2017). Bivalent chromatin comprises of an active chromatin mark H3K4me3 and a repressive H3K27me3 and is protected from DNA methylation in somatic cells (Bernstein et al., 2006; Brinkman et al., 2012). DNMT3A1, the longer

isoform containing the intrinsically disordered N-term domain, predominates in adult somatic tissues and generally follows the dynamics of the H3K27me3 mark during neuronal differentiation, although these two marks do not tend to overlap at CGIs (Brinkman et al., 2012; Manzo et al., 2017).

1.11.2 The ADD domain and H3K4me3

The ADD domain, present in all DNMT3s, is homologous to a conserved PHD (plant-homeodomain) zinc finger motif. ADD domains of DNMT3s have a high affinity to the N-terminal region of histone 3, which is ablated in presence of H3K4me2/3, H3K4ac, H3T3ph, H3S10ph, and/or H3T11ph PTMs (Zhang et al., 2010). Methylated lysine H3K4me3 in somatic cells is found at active gene promoters and TSSs, and inhibits DNMT3 activity (Li et al., 2011; Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010). Structural studies have found that DNMT3A is intrinsically in an autoinhibitory allosteric conformation, driven by the ADD domain: the ADD domain masks the DNA binding site of the catalytic domain. Recognition of unmethylated H3K4 specifically allows a structural shift and uncouples the ADD-catalytic domain interaction, allowing activation of DNMT3A enzymatic function (Figure 1.4a) (Guo et al., 2015; Li et al., 2011; Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010). Similarly, recruitment of catalytically inactive DNMT3L ADD domain by unmethylated H3K4 is sufficient to engage the whole complex, for example, as may occur in the case of DNMT3L-DNMT3A in the oocyte (Jia et al., 2007; Ooi et al., 2007). As a result, DNA methylation and H3K4me3 are mutually exclusive in the genome and H3K4me3 protects promoters from silencing by DNA methylation. Engineering of the ADD domain to lose sensitivity to H3K4me3 or H3T5ph results in aberrant gain of methylation over these domains and ESC differentiation failure or chromosomal instability, respectively (Noh et al., 2015).

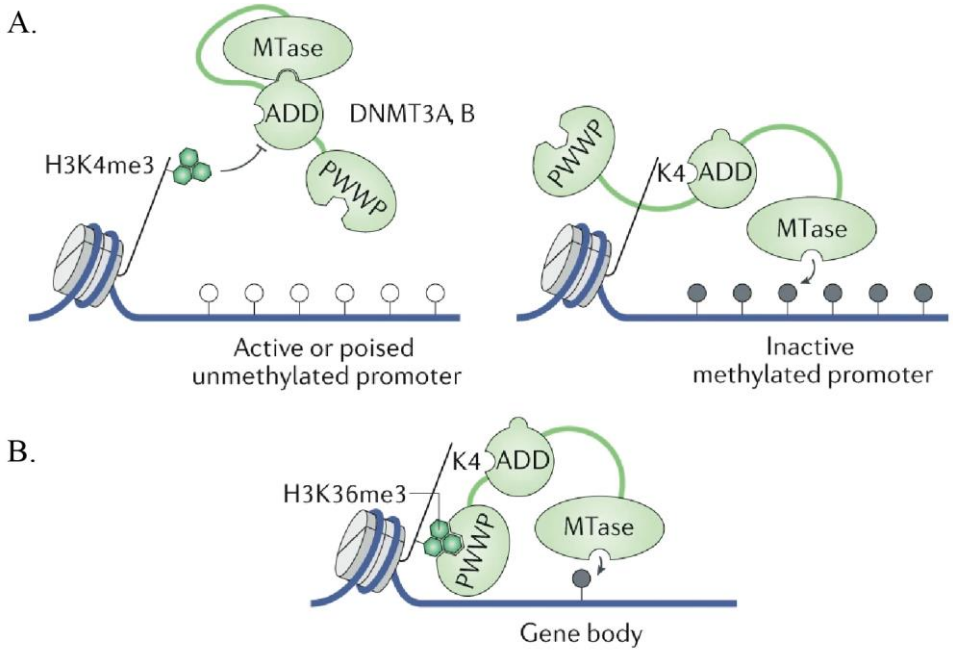


Figure 1.4. DNMT3A/B recruitment mechanisms.H3K4me3, found at promoters of active or bivalent (poised) promoters is recognised by the ADD domain, and DNMT3A/B remain in an autoinhibitory state. When H3K4 site in unmethylated, e.g. at inactive promoters, ADD domain is recruited to this PTM. Engagement of ADD domain leads to a change of DNMT3A/B conformation and result in DNMT3 catalytic activity. **B.** H3K36me3 mark, found at gene bodies, together with unmethylated H3K4 are recognised by the PWWP and ADD domains, respectively, and lead to DNMT3A/B activation. Image reproduced from (Greenberg and Bourc'his, 2019).

1.11.3 The PWWP domain and H3K36me3

The PWWP domain is a member of the Tudor domain royal superfamily and is mostly found in chromatin-interacting proteins. It has an intrinsic and somewhat unspecific affinity to chromatin and modified histones, especially lysine residues, via a conserved aromatic cage (Qin and Min, 2014; Wu et al., 2011). In the DNMT family, the PWWP domain is only found in DNMT3A and DNMT3B. The PWWP domain is responsible for non-specific DNA and chromatin binding of DNMT3s (Ge et al., 2004; Qiu et al., 2002) and is required for methylation of major satellite repeats at pericentromeric chromatin (Chen et al., 2004).

Extensive biochemical, structural and modelling work suggests that the PWWP domain of DNMT3A and DNMT3B interacts specifically and exclusively with H3K36me2/3 (Bock et al., 2011; Dhayalan et al., 2010; Dukatz et al., 2019; Mauser et al., 2017; Rondelet et al., 2016; Weinberg et al., 2019). H3K36me3 is universally found over expressed gene bodies, and in mouse ESCs the PWWP domain of DNMT3B is required for DNA methylation over H3K36me3 domains (Baubec et al., 2015; Neri et al., 2017). This interaction appears to be specific to the PWWP domain of DNMT3B but not DNMT3A (Baubec et al., 2015; Morselli et al., 2015). ADD and PWWP domains can work in a synergistic manner to recognise a combination of unmethylated H3K4 and H3K36me3 on histone tails (Figure 1.4b) (Gong et al., 2020).

1.12 Generation of a PWWP-DNMT3A mutant mouse model

In vitro studies show that the PWWP domain of DNMT3A interacts with H3K36me2/3 but a point mutation (D329A) in a conserved aspartic acid residue in the PWWP domain abrogated the ability of the DNMT3A-PWWP domain to bind H3K36me2 and H3K36me3 peptides or pull down H3K36me2/3-containing native nucleosomes (Dhayalan et al., 2010). The D329A mutation also reduced the preferential heterochromatin localisation of DNMT3A2 in transfected cells and decreased catalytic activity on unmethylated native nucleosomes. While the PWWP domain is responsible for DNMT3B gene-body targeting, the function role of the PWWW domain in DNMT3A recruitment to the genome *in vivo* remained unknown. A mouse model carrying the previously described D329A mutation in *Dnmt3a* was generated to investigate the *in vivo* significance of the DNMT3A PWWP domain.

The D329A mutation was introduced in the endogenous *Dnmt3a* exon 8 locus using homologous recombination in mouse ESCs (Figure 1.5) (Sendžikaitė et al., 2019). The targeting vector with a point mutation flanked by homology arms and selection cassettes was transfected into black-coated C57BL/6 ESCs. After confirmation of successful recombination, ESCs were injected into albino C57BL/6 recipient blastocysts and re-implanted into pseudo-pregnant females. Offspring showing high chimerism rate were used for the breeding with mice carrying constitutively expressed *Cre* recombinase, to remove the positive selection cassette.

Litters with paternal transmission of mutant $Dnmt3a^{D329A}$ showed over 50% post-weaning survival, whilst there were no survivors when the mutant allele was transmitted

maternally. Therefore, *Dnmt3a*^{D329A} males were crossed to females with heterozygous oocyte-specific conditional ablation of *Dnmt3a* (*Δ*) (Kaneda et al., 2010, 2004), generating four genotypes amongst offspring: *+/+*, *Δ/+*, *+/D329A* and *Δ/D329A* (where the maternal allele is listed first); the last carries the D329A mutation as the only functional *Dnmt3a* allele (Figure 1.6).

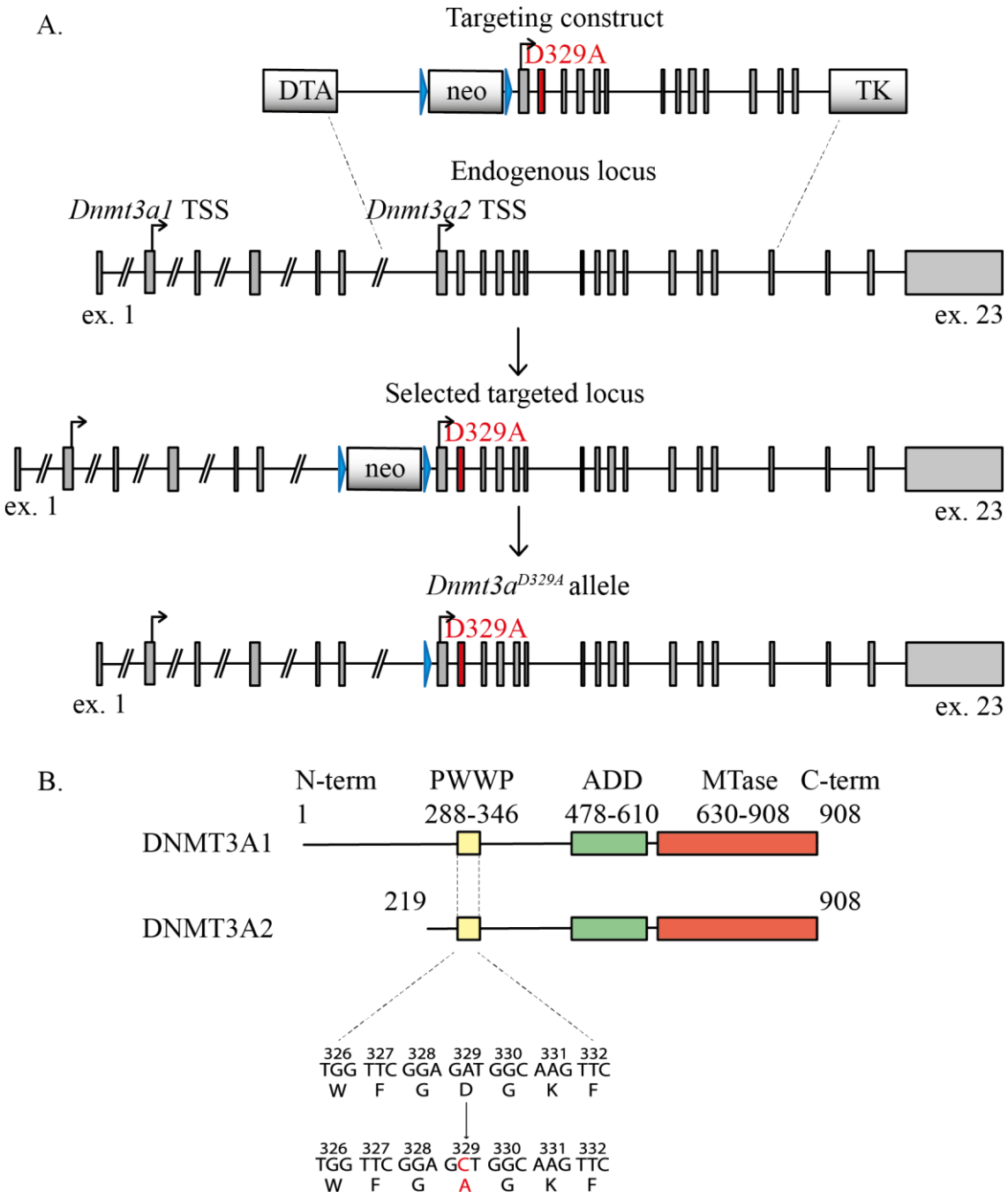


Figure 1.5. Introduction of the D329A mutation at endogenous *Dnmt3a* locus. A. Schematic representation of targeting construct used to generate mice carrying the

Dnmt3a^{D329A} mutation. Targeting construct, endogenous locus and selected targeted locus are shown together with final neo cassette-excised mutant allele. Exon 8 with the D329A mutation is coloured in red; *loxP* sites indicated as blue arrowheads. Abbreviations: DTA: diptheria toxin A negative selection cassette, neo: neomycin selection cassette, TK: herpes simplex virus thymidine kinase, TSS: transcription start site, ex.: exon. **B.** Schematic representation of the DNMT3A protein isoforms and location of the base substitution, coloured red, introduced to generate the D329A mutation. NCBI IDs: CCDS36397.1, CCDS25784.1. Known domains are indicated by coloured boxes, based on the PROSITE database. Numbers indicate amino acids.

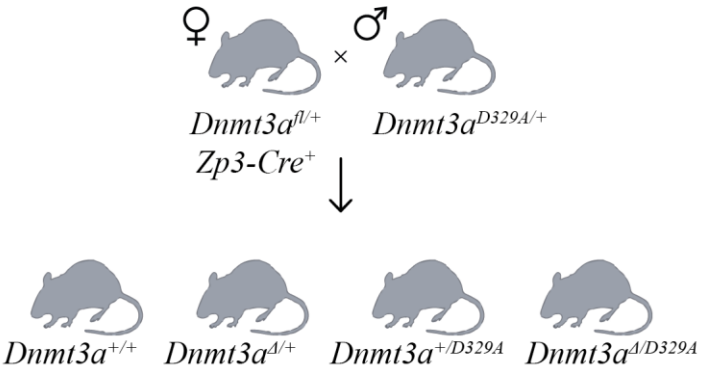


Figure 1.6. Transgenic mice breeding scheme. Breeding scheme used to generate offspring of four different genotypes. *fl*: *Dnmt3a* containing *loxP* sites surrounding exon 18; *Δ*: *loxP* sites after recombination leading to deletion of exon 18; *Zp3-Cre*: oocyte-specific *Zp3* promoter-driven *Cre* recombinase (Kaneda et al., 2010, 2004); *D329A*: a *Dnmt3a* allele point missense mutation in exon 8.

1.13 Aims

The work outlined below aims to deepen the understanding of methyltransferase regulatory domain roles in DNMT targeting to the genome. In particular, it aims to understand the functional role of the DNMT3A PWWP domain *in vivo* by use of a mouse model carrying the D329A mutation, previously shown to be important in DNMT3A-PWWP and H3K36me3 interaction (Dhayalan et al., 2010).

First, the oocyte was utilised as a system where DNMT3B does not play a role and strong correlation between DNA methylation and H3K36me3 is found. It was initially hypothesised that introduction of the D329A mutation would result in loss of DNA methylation in the affected oocytes; this work is described in Chapter 3. After the serendipitous observations of phenotypic changes and rejection of the initial hypothesis in the oocyte, further aims were set to understand the phenotypic consequences of the presence of the *Dnmt3a*^{D329A} mutation in mice. Brain is another tissue, apart from oocyte, where DNMT3A is a dominant *de novo* methyltransferase. Therefore, I aimed to utilise this tissue to describe the molecular changes in DNA methylation patterns introduced by DNMT3A^{D329A}, altered transcriptional and chromatin environment and protein interactions of mutant DNMT3A^{D329A}.

2 Methods and materials

2.1 Animal experimental procedures and sample collection

All animal experimental procedures were approved by the Animal Welfare and Ethical Review Body at the Babraham Institute, and were conducted under authority of the UK Home Office issued licenses in accordance with the Animal (Scientific Procedures) Act 1986.

The *Dnmt3a*^{D329A} mutant strain was generated by GenOway (France) by targeting a point mutation GAT to GCT resulting in aspartic acid to alanine amino acid change at the endogenous locus of *Dnmt3a* codon 329 (Sendžikaitė et al., 2019). These mice were inter-crossed with C57BL/6Babr mice carrying a conditional *Dnmt3a* deletion allele *Dnmt3a*^{fl} and oocyte-specific *Zp3* promoter driven *Cre* recombinase (Kaneda et al., 2010). In order to obtain experimental litters *Dnmt3a*$^{fl/+}$ Zp3-*Cre*$^{+ve}$ females, which carry oocytes with a conditionally deleted *Dnmt3a*$^{\Delta}$ allele, were crossed with *Dnmt3a*$^{D329A/+}$ males; yielding litters with four possible genotypes: *Dnmt3a*$^{+/+}$, *Dnmt3a*$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$, and *Dnmt3a*$^{\Delta/D329A}$ (Figure 1.6).

When possible, body weights of experimental litters were recorded at 1.4, 4, 6, 8, 10, and 12 weeks and analysed using mixed-model ANOVA with *post-hoc* pairwise comparisons using Bonferroni correction.

Litters were sacrificed and tissues were collected at embryonic day 7.5 (E7.5), postnatal day 1 (P1), P25, and adulthood (~14 weeks old). At E7.5 ectoplacental cone (EPC) and epiblast were collected in PBS and EZ Nuclear Lysis buffer (Sigma), respectively.

2.2 Oocyte collection

For collection of bulk oocyte samples, dissected ovaries were submerged in digestion buffer (2mg/ml collagenase (Sigma-Aldrich) and 0.025% Trypsin (Sigma-Aldrich) diluted in PBS) and put into the thermomixer at 37°C for 30min incubation. The ovaries were manually dissociated by short repetitive pipetting after 5, 15 and 25min. After digestion, dissociated tissue suspension was mixed with M2 medium (Sigma-Aldrich) in

1:1 ratio. Oocytes were then collected using a mouth-controlled glass pipette and snap-frozen on dry ice in PBS.

For single cell oocyte samples, ovaries in M2 medium were scraped using a sterile needle to release oocytes. These were then collected using a mouth-controlled glass pipette and snap-frozen on dry ice in RLT+ buffer (QIAGEN).

2.3 Genotyping

E7.5 EPCs were lysed in embryo lysis buffer (1M Tris pH8.5, 0.5M EDTA pH8, 10% Tween, 0.2mg/ml Proteinase K) and used for genotyping by PCR with primers for *Dnmt3a*$^{+/fl/\Delta}$ alleles (D3A-F 5'-CTGTGGCATC TCAGGGTGAT GAGCA-3', D3A-R1 5'-GCAAACAGAC CCAACATGGA ACCCT-3', and D3A-R2 5'-TGAGTGGTGA GGCCCAGCTT ATCGA-3') and *Dnmt3a*^{D329A} allele (D329A-F 5'-CAGATCCTTG CCTGAACTGT GGTGC-3' and D329A-R 5'-TCCCTCTTGG TCCAGCATGT ACCCT-3'). Adult mice were genotyped by the same PCR assay using genomic DNA extracted from ear clips, or by service provider Transnetyx (US) using real-time PCR.

2.4 ELISA

Blood samples from experimental animal hearts were collected into Microvette 500 Z-Gel tubes (Sartstedt), centrifuged for 5 min at 10,000 g at RT. Serum collected was used in Rat/Mouse Growth Hormone ELISA kit (Millipore) and Quantikine ELISA for Mouse/Rat Igf-1 kit (R&D systems) as per manufacturers' instructions.

2.5 Post-bisulphite adaptor tagging

Post-bisulphite adaptor tagging (PBAT) was used to generate genome-wide DNA methylation profiles. Approximately 50ng of genomic DNA or 10% of MNase-digested E7.5 Epiblast DNA, purified using solid phase reversible immobilization (SPRI) beads, was used as an input material. For bulk oocyte samples, approx. 100-oocyte sample in PBS was used as an input.

If needed, DNA was extracted using DNeasy blood and tissue kit (QIAGEN), as per manufacturers' instructions. Imprint DNA modification kit (MOD50, Sigma) one-step procedure was used to bisulphite-convert the DNA. Sample volumes were adjusted with Elution Buffer (EB) (Qiagen) to final volume of 25µl and 125µl of bisulphite conversion

reagents from the kit were added. Samples were incubated at 99°C for 6min and at 60°C for 90mins. Samples were cleaned using Sigma Imprint columns and eluted in 20µl of EB. First strand synthesis was done by addition of 0.4mM dNTP mix (Invitrogen), 10x NEB buffer 2 (B7002, NEB), 0.4µM Illumina adaptor and 9 random nucleotide 9N31F primer. Samples were incubated at 65°C for 3mins and cooled down to 4°C. Five units of Klenow exo- (M0212, NEB) were added and samples were incubated at 4°C for 5mins, then temperature was increased from 4°C to 37°C in 1°C per 15sec steps. Then samples were incubated at 37°C for 90mins. This was followed by addition of 20 units of Exonuclease-I (M0293, NEB) and incubation at 37°C for 60mins. DNA was purified using 0.8:1 ratio SPRI beads.

For biotinylated DNA capture, 20µl of Dynabeads Strepatavidin M-280 beads (11205, ThermoFisher Scientific) per sample were washed twice in Bind and Wash buffer (10mM Tris-HCl pH7.5, 1mM EDTA, 2M NaCl) and resuspended in 50µl volume per sample. DNA from first strand synthesis was added to washed streptavidin beads and incubated for 20min at RT with overhead rotation. Afterwards the supernatant was discarded and beads were washed twice in 0.1 N NaOH and then twice in EB. Second strand synthesis was assembled by addition of 0.4mM dNTP mix, 10x NEB buffer 2, 0.4µM 9N31R primer and water to final volume of 48µl. Samples were incubated at 95°C for 1min and cooled down to 4°C. Ten units of Klenow exo- (M0212, NEB) were added and samples were incubated at 4°C for 5mins, then temperature was increased from 4°C to 37°C in 1°C per 15sec steps. Then samples were incubated at 37°C for 90mins.

The beads were washed twice in EB buffer and resuspended in library amplification mix, containing 0.2mM dNTPs, 5x KAPA HiFi buffer (Kapa Biosystems), 1 unit KAPA HiFi polymerase, 0.2µM of Illumina forward primer P1.0 and 0.2µM selected Illumina multiplex reverse oligonucleotide and water to the final volume of 50µl. Sample DNA was initially denatured at 95°C for 2mins, followed by 10 cycles of 80sec at 98°C, 30sec at 60°C and 30sec at 72°C; with final extension of 3mins at 72°C. Libraries were purified using 0.8:1 ration SPRI beads. Libraries were quantitated and quality control checked using Agilent High Sensitivity DNA kit and Kapa Illumina library quantitation kit, as per manufacturers' instructions. Libraries were sequenced using 100bp single-end for oocyte, P1, P25, adult tissues, and ESCs; and 150bp for E7.5 epiblast sequencing mode on the Illumina NextSeq500 platform at the Babraham Institute Sequencing Facility.

2.6 Single-cell bisulphite sequencing

Single cells in 10µl of RTL+ buffer were used as an input. EZ DNA methylation direct kit (D5021, Zymo) was used to bisulphite-convert the DNA. First, the sample was incubated in 65µl of CT conversion Reagent for 8min in 98°C, then at 64° for 3 hours. Bisulphite converted DNA was purified following EZ DNA methylation direct kit instructions. First strand synthesis was done by addition of 0.4mM dNTP mix (Invitrogen), 10x blue buffer (B0110, enzymatics), 0.4µM adaptor 6 random nucleotide 6NF primer and water to the final volume of 49µl. Samples were incubated at 65°C for 3mins and cooled down to 4°C. 50 units of Klenow exo- (P7010-HC-L, Enzymatics) were added and samples were incubated at 4°C for 5mins, then temperature was increased from 4°C to 37°C in 1°C per 15sec steps. Then samples were incubated at 37°C for 30mins.

Samples were incubated at 95°C for 45secs and cooled down. Additional 2.5µl of first strand synthesis master mix were added and the same thermocycling protocol was followed. These steps were repeated 4 times in total. This was followed by addition of 40 units of Exonuclease-I and incubation at 37°C for 60mins. DNA was purified using 0.8:1 ratio AMPureXP beads (Beckman), leaving the beads in the sample for the next steps.

Second strand synthesis was assembled by addition of 0.4mM dNTP mix, 10x blue buffer, 0.4µM 6NR primer and water to final volume of 49µl. Samples were incubated at 98°C for 2min and cooled down to 4°C. 50 units of high concentration Klenow exo- were added and samples were incubated at 4°C for 5mins, then temperature was increased from 4°C to 37°C in 1°C per 15sec steps. Then samples were incubated at 37°C for 90mins.

DNA was then purified by addition of AMPureXP buffer only to complement the beads present in the sample. Beads were resuspended in library amplification mix, containing 0.2mM dNTPs, 5x KAPA HiFi buffer (Kapa Biosystems), 1 unit KAPA HiFi polymerase, 0.2µM of Illumina forward primer P1.0 and 0.2µM selected Illumina multiplex reverse oligonucleotide and water to the final volume of 50µl. Sample DNA was initially denatured at 95°C for 2mins, followed by 14 cycles of 80sec at 94°C, 30sec at 65°C and 30sec at 72°C; with final extension of 3mins at 72°C. Libraries were purified using 0.8:1 ratio of AMPure buffer and beads already present in the sample. Libraries were quantitated and quality control checked using Agilent High Sensitivity DNA kit and Kapa Illumina library quantitation kit, as per manufacturers' instructions. Libraries were

sequenced using 75bp single-end sequencing mode on the Illumina NextSeq500 platform at the Babraham Institute Sequencing Facility.

2.7 RNA sequencing

Samples were processed by Elena Ivanova.

Total RNA from 14-week old animal hypothalami was extracted by submerging in TRIsure (Meridian bioscience) and dissociating the tissue using electric homogenizer. Samples were incubated at RT for 5mins, centrifuged at 12,000 g for 15mins at 4°C and the aqueous phase was taken for RNA extraction. Cold isopropyl alcohol in ratio 1:2 was added to the samples and incubated for 10mins at RT. Samples were centrifuged at 12,000 g for 10mins at 4°C and the supernatant was removed. Pellets were washed once with 75% ethanol in ratio 1:1 to initial TRIsure used. Samples were centrifuged at 7,500 g for 5mins at 4°C. Pellets were air-dried and dissolved in water.

RNA was cleaned using Ribo-Zero rRNA removal kit (Illumina), where DNA probes to rRNA are hybridised and selectively cleaved using RNA-DNA hybrid targeting RNAse. Libraries were generated using NEBNext Ultra RNA library preparation kit (E7420, New England Biolabs) as per manufacturers' instructions. Briefly, RNA was fragmented by incubation at 94°C for 15mins. cDNA was synthesised using ProtoScript II reverse transcriptase and random primers, the first and second strands were synthesised using reagents provided in the kit and purified using AMPure XP Beads. The ends in cDNA library were repaired and NEBNext adaptors for Illumina were blunt ligated. The ligation reaction was purified using AMPure XP beads and library was amplified using NEBNext Q5 polymerase, USER enzyme, and Illumina sequencing index primers. Libraries were purified using 0.9:1 ration of AMPure buffer and beads already present in the sample. Libraries were quantitated and quality control checked using Agilent High Sensitivity DNA kit and Kapa Illumina library quantitation kit, as per manufacturers' instructions. RNA-seq libraries were sequenced 50bp single-end using Illumina HiSeq2500 platform.

2.8 Ultra-low input chromatin immunoprecipitation sequencing

Ultra-low input native chromatin immunoprecipitation sequencing (ChIP-seq) libraries were generated using 2.5% of a whole hypothalamus as input for each immunoprecipitation, following previously described protocol (Hanna et al., 2018), keeping 10% of that as input. Cells were permeabilised by adding 1/10 volume of 1%

triton X-100 and 1% deoxycholate solution. Chromatin was then digested using 2U/µl micrococcal nuclease (M0247, NEB) in 10x NEB MNse buffer, 1mM DTT, 10% PEG6000. Samples were digested for 7.5mins at 21°C. A solution of 100mM EDTA, 1% triton, 1% deoxycholate was added in ratio 1:10 to stop the reaction and samples were rested on ice for 10mins. Each sample had 150µl of complete immunoprecipitation buffer (20mM Tris-HCl, pH8.0, 2mM EDTA, 150mM NaCl, 0.1% Triton-X100, 1x protease inhibitor cocktail, 1mM PMSF) added and samples were incubated at 4°C rotating for 1hr. Afterwards, input fraction was taken for all samples.

For antibody binding, two sets of 5µl of Dynabeads Protein A/G beads (10002D and 10004D, ThermoFisher Scientific) per sample were washed three times in complete immunoprecipitation buffer. Antibodies for H3K4me3 (250ng per reaction, Diagenode C15410003), H3K27me3 (125ng per reaction, Millipore 07-449), or H3K36me3 (250ng per reaction, Diagenode C15410192) in complete immunoprecipitation buffer were added to one aliquot of Protein A/G beads per sample and incubated at 4°C for 3hrs. Meanwhile, the samples were precleared by adding the other 5µl of washed Protein A/G beads to each sample and rotating at least for 2hrs at 4°C.

The supernatant from antibody-bound beads was discarded, and the pre-cleared chromatin was transferred to antibody bound Protein A/G for incubation at 4°C overnight with rotation. The next day beads were washed twice with low salt buffer (20mM Tris-HCl, pH8.0, 2mM EDTA, 150mM NaCl, 1% Triton-X100, 0.2% SDS) and then twice with high salt buffer (20mM Tris-HCl, pH8.0, 2mM EDTA, 500mM NaCl, 1% Triton-X100, 0.2% SDS). Bound DNA was eluted by adding 30µl of ChIP elution buffer (100mM $NaHCO_3$, 1% SDS) and incubating for 1.5hrs at 65°C. DNA was purified by 1.8:1 ratio SPRI bead purification. MicroPlex Library Preparation kit v2 (Diagenode) was used as per manufacturer's instructions to generate libraries of inputs and immunoprecipitated samples. Libraries were quantitated and quality control checked using Agilent High Sensitivity DNA kit and Kapa Illumina library quantitation kit, as per manufacturers' instructions. Libraries were sequenced using 75bp single-end sequencing on the Illumina NextSeq500 platform at the Babraham Institute Sequencing Facility.

2.9 Cloning

Plasmids pJET-Dnmt3a1, containing full *Dnmt3a1* sequence, and mammalian expression vector pCAGbio-eGFP, containing *loxP*-flanked CAG promoter, *AviTag* sequence, and

eGFP, were kindly provided by T. Baubec (University of Zurich). A point mutation was introduced using QuickChange II site directed mutagenesis kit (Agilent) using primers (Fwd 5'-GTCATGTGGTTCGGAGCTGGCAAGTTCTCAGTG-3' and Rev 5'-CACTGAGAACTTGCCAGCTCCGAACCACATGAC-3') and confirmed using Sanger sequencing. pCAGbio-eGFP and pJET-Dnmt3a1^{D329A} was cut using Sal I and Not I enzymes (NEB), run on a 1% agarose gel and purified using Monarch DNA gel extraction kit (NEB). Backbone fragment of pCAGbio and insert fragment of Dnmt3a1^{D329A} were then ligated using T4 DNA ligase (NEB) and transfected into DH5α competent cells for subcloning (Thermo Scientific), all as per manufacturers' instructions. Bacterial cells were streaked on LB agar plates with 100 µg/ml ampicillin and left overnight in a 37°C incubator. The next day, colonies were picked from the plates and grown overnight in liquid LB and ampicillin (100 µg/ml) culture. A miniprep purification of plasmid was done using QIAprep spin miniprep kit (QIAGEN).

2.10 ESC culture

HA36CB1 cell line derived from 129xC57BL/6 mESCs with constitutively expressed biotin-ligase (BirA) and recombinase-mediated cassette exchange (RMCE) construct (referred to as *wild-type*) and a variant with *AviTag* marked *Dnmt3a1* were used in experiments (Lienert et al., 2011). They were kindly provided by Tuncay Baubec (University of Zurich).

Cells were cultured in serum-containing ESC medium at 37°C and 5% CO_2 incubator on 0.2% gelatin-coated tissue culture dishes. ESC medium was composed of DMEM (High glucose, pyruvate, no glutamine, Invitrogen), 15% foetal bovine serum (Invitrogen), 100µM non-essential amino acids (Invitrogen), 2mM L-Glutamine (Invitrogen), 0.01% β-mercaptoethanol (Sigma-Aldrich) and a titrated vial of home-made leukemia inhibitory factor (LIF) (provided by T. Baubec, University of Zurich).

The medium was changed every day and cells were passaged by removing the medium, washing with PBS (Invitrogen), then incubating with 0.05% Trypsin-EDTA (Invitrogen) for 5min at 37°C and quenching the reaction by addition of DMEM with 9% foetal growth serum.

2.11 Neural progenitor differentiation

ESCs were counted and 3.5 million cells were plated on a non-adhesive bacterial dish in CA medium (DMEM, 10% foetal bovine serum, 100mM non-essential amino acids, 2mM L-Glutamine, 0.01 β-mercaptoethanol). On day 2 medium was changed and cells were replated into new dishes. On day 4 medium change was performed, with addition of 5µM retinoic acid. On day 6 medium was changed again adding retinoic acid. Cells were harvested on day 8.

2.12 Cell line generation

2.12.1 Preparation of plasmid DNA prior electroporation

Plasmids pIC-Cre (Gu et al., 1993) and pCAGbio-Dnmt3a1-D329A, as described in Methods section 2.9, were precipitated by addition of 0.3M sodium acetate (Sigma Aldrich) and 75% ethanol (Fisher Scientific) overnight at -20°C. DNA was centrifuged at 4°C at max speed of table-top centrifuge for approx. 2hr, then washed with 70% ethanol and centrifuged for another 30min at 4°C at max speed. DNA was air dried under the hood.

2.12.2 Electroporation

The cells were detached as per passage protocol and counted to get $2x10^6$ cells for each transfection. Cells were centrifuged at 200 g for 10mins and after removal of supernatant mixed with 100µl of supplemented Amaxa solution (Lonza). Solution was then combined with 15µg pIC-Cre and 25µg of target plasmid, transferred to Amaxa cuvette, and placed into the Nucleofector (Lonza) for electroporation under mESC programme. Afterwards, cell solution was diluted in ESC medium and plated in two different concentrations. After 24hrs, one media change was done. On day 2 the selection was started by addition of 3µM Ganciclovir. Subsequently media was changed every 2-4 days depending on the dead cell numbers, always adding 3µM Ganciclovir.

2.12.3 DNA extraction

After 12 days individual colonies were picked and replated individually, splitting each colony into two. One part was left for expansion and the other copy was used for PCR genotyping to confirm insertion. On day 14 cells for genotyping were lysed in lysis buffer (0.025% SDS, 0.3 KCl, 5mM EDTA, 25 mM TRIS pH 8.3, 1% NP-40, 2% Tween-20,

and Proteinase K 1mg/ml (Roche). Cells were incubated at 55°C for 1hr, then proteinase K was inactivated by incubation at 95°C for 10mins.

2.12.4 Nuclear protein extraction

Cell were pelleted for 10mins at 200 g and then resuspended in Buffer A (0.01M EDTA, 0.01M KCl, 0.01M HEPES, 1mM DTT, 1x protease inhibitor cocktail, 0.1% NP-40) and incubated for 10mins at 4°C. Cells were then centrifuged for 10mins at 4°C at 2000 g and the cytosolic fraction was removed. Pelleted nuclei were resuspended in Buffer B (17mM HEPES, 0.85% glycerol, 0.85mM EDTA, 1mM DTT, 1x protease inhibitor cocktail) and incubated in a thermomixer for 1hr at 4°C at 2000rpm. Samples were then centrifuged in a tabletop centrifuge at max speed for 10mins at 4°C. The supernatant had nuclear protein fraction. Protein concentration was quantified using Qubit protein assay kit (Q33211, Qubit), as per manufacturer's instructions.

2.12.5 Western blotting

20μg of nuclear lysate was mixed with Laemmli buffer (2% SDS, 5% β-mercaptoethanol, 20% glycerol, 0.01M Tris pH 6.8, 0.005% bromophenol blue) and boiled at 95°C for 10mins. This mixture was loaded in a well of an SDS-PAGE gel with 8% resolving gel (46% water, 8.8% acrylamide, 0.39M Tris pH 8.8, 0.1% SDS, 0.1% APS, 0.06% TEMED) and 5% stacking gel (70% water, 5.5% acrylamide 0.2M Tris pH 6.8, 0.1% SDS, 0.1% APS, 0.1% TEMED). The gel was run in SDS Running buffer (25mM Tris, 0.19M Glycine, 1% SDS) at 120V.

The PVDF membrane for western blot (Millipore) was first activated in 100% methanol for 3mins and then washed with water. Sponges, Whatman paper, gel and membrane were soaked in transfer buffer (50mM Tris, 0.4M Glycine, 0.1% SDS, 20% methanol) and proteins were transferred from the gel onto the membrane for 1hr at 100V in the cold room. The membrane was washed twice with TBST buffer (01M Tris pH 8, 0.15M NaCl, 0.1% Tween-20) and incubated at room temperature for 1hr with 5% milk in TBST solution. Afterwards, membrane was incubated on a shaker overnight at 4°C with a 2mg/ml primary antibody for DNMT3A1 (ab2850, Abcam) diluted in TBST. Next day the membrane was washed in TBST three times for 10mins and incubated with a 0.2mg/ml secondary goat anti-rabbit IgG (H+L) HRP antibody (31466, Invitrogen) in TBST for 1hr at room temperature. After another three 10min washes with TBST the

membrane was incubated in Amersham ECL Western Blotting Detection Reagent (RPN2109, GE Healthcare Life Sciences) for 1min and then developed on Amersham Hyperfilm ECL (28906836, GE Healthcare Life Sciences).

For detection of biotinylated proteins, membrane was incubated in Restore Western Blot Stripping Buffer (21059, Thermo Fischer Scientific) for 5mins at room temperature. The membrane was then washed three times for 10mins in TBS-TX (TBS with 0.1% Triton X-100). The membrane was blocked with 4% BSA in TBS-TX for 1hr at room temperature. The membrane was incubated with 0.1mg/ml Pierce high sensitivity streptavidin-HRP antibody (21130, Thermo Scientific) for 1hr at room temperature and washed twice for 10mins in TBS with 0.3% Triton X-100. Afterwards the membrane was washed twice for 10mins in TBS with 0.3% Triton X-100 and 0.5M NaCl and finally washed once for 10mins in TBS with 0.3% Triton X-100 before detection of protein using ECL, as described in previous paragraph.

2.13 Biotin co-immunoprecipitation for Mass Spectrometry

Samples were processed together with Joel Wirz (University of Zurich).

2.13.1 Nuclear extract preparation

Cells were harvested by detaching with 0.05% Trypsin-EDTA for 5min at 37°C and quenching the reaction by addition of DMEM with 9% foetal growth serum. Cells were then washed in DPBS and resuspended in nuclear extract buffer 1 (10mM HEPES pH 7.5, 10mM KCl, 1mM EDTA, 1.5 mM $MgCl_2$, 1mM DTT, 1 x protease inhibitor cocktail) and incubated for 10mins at 4°C. Cells were then centrifuged for 10mins at 4°C at 2,000 g and supernatant was removed. Pelleted nuclei were resuspended in nuclear extract buffer 2 (20mM HEPES pH 7.5, 20% glycerol, 1.5mM $MgCl_2$, 0.2mM EDTA, 1mM DTT, 1 x protease inhibitor cocktail, 450mM NaCl) and transferred to a Dounce Homogenizer. Ten strokes were done using a loose pestle, then the suspension was transferred to a fresh tube and centrifuged at 2500 g for 10mins. The supernatant was removed, nuclei pellet resuspended in nuclear extract buffer 1 and 300units/ml benzonase and incubated rotating for 3hrs at 4°C. Samples were centrifuged at 2500 g for 10mins and resuspended in nuclear extract buffer 2 (20mM HEPES pH 7.5, 20% glycerol, 1.5mM $MgCl_2$, 0.2mM EDTA, 1mM DTT, 1x protease inhibitor cocktail) and 450mM NaCl. Samples were transferred to Dounce homogeniser and 10 strokes were done using a tight

pestle. Samples were then incubated on a shaker at 4°C for 1hr to extract nuclear proteins. Afterwards, they were centrifuged at 16,000 g at 4°C for 15mins, the supernatant fraction was kept.

2.13.2 Streptavidin immunoprecipitation

Dynabeads M-280 Streptavidin magnetic beads were washed with IP buffer (20mM HEPES pH 7.5, 20% glycerol, 1.5mM $MgCl_2$, 0.2mM EDTA, 1mM DTT, 1 x protease inhibitor cocktail, 150mM NaCl, 0.3% NP40) three times for 10mins at 4°C. Beads were pre-blocked in IP buffer with addition of 1% Fish skin gelatin for 1 hr at 4°C.

The samples were diluted 1:2 ratio in nuclear extract buffer 2 without NaCl, to achieve final NaCl concentration of 150mM. Then samples were centrifuged for 15 min at 16,000 g at 4°C and supernatant was kept. NP-40 was added to final concentration of 0.3% and protein concentration was quantified using Qubit protein assay, as per manufacturer's instructions. Approx. 1mg of nuclear extract was used for one IP, 5% of it kept as input. Pre-blocked streptavidin beads were added and samples were incubated overnight with overhead rotation at 4°C. The next day beads were washed once rotating for 10mins at 4°C in wash buffer (20 mM HEPES pH 7.5, 20% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 150 mM NaCl, 0.3% NP-40, 1 mM DTT, and 1× protease inhibitor cocktail) and then again in the same buffer without NP-40. The beads were washed for 10 min in 50 mM Tris pH 8.5, 1 mM EGTA, 75 mM KCl, rotating overhead at 4°C.

Proteins were eluted by adding 200 µl of UA solution (8 M urea (GEPURE00-66) in 100 mM Tris–HCl pH 8.2) with 0.1 M DTT and shaking for 30 min at room temperature in a thermomixer. Processing for label-free mass spectrometry was done immediately afterwards.

2.14 Label-free mass spectrometry

The whole supernatant from streptavidin IP was digested using filter-aided sample preparation (FASP) method, as per published protocol (Wiśniewski et al., 2009). Samples were loaded to the filter unit and centrifuged at 14,000 g for 25mins at RT. This centrifugation setting was used for all subsequent steps, unless indicated otherwise. 200µl UA solution was added and samples were centrifuged. 100µl of 0.05M Iodocetamine in UA solution was added, samples were incubated in a thermomixer at 600rpm for 1min, then incubated for 5mins on the bench and centrifuged. 100µl of UA solution was added

and samples were centrifuged, this step was repeated three times in total. 100µl of 0.5M NaCl were added to the samples and centrifuged, this step was repeated twice. 120µl of 0.05M Triethylammonium bicarbonate with 20ng/µl trypsin were added and samples were incubated in a thermomixer at 600rpm for 1min, then overnight on the bench. The next day samples were centrifuged and supernatant collected. The samples were acidified by addition of trifluoroacetic acid to the final concentration of 0.5%.

Samples were then desalted with C18 StageTips, made in house at Functional Genomics Centre Zurich. StageTips were activated by addition of 145µl 100% methanol, and centrifuged for 1min at 2,000 g. This centrifuge speed was used for subsequent centrifugations. StageTips were cleaned by adding 145µl of 60% acetonitrile and 0.1% trifluoroacetic acid followed by centrifugation for 1min. StageTips were equilibrated by adding 145µl of 3% acetonitrile and 0.1% trifluoroacetic acid, followed by centrifugation for 1min. This step was done twice in total.

Digested protein samples were adjusted by adding acetonitrile to the final concentration of 3%. Samples were loaded 145µl at a time into the StageTips, followed by centrifugation for 1.5min. The flow-through was collected and loaded to the StageTips again, followed by centrifugation for 1.5min. The peptides were desalted by adding 145µl of 3% acetonitrile and 0.1% trifluoroacetic acid and centrifugation for 1min, repeated twice in total. Peptides were eluted using 60% acetonitrile and 0.1% trifluoroacetic acid and centrifugation for 2mins, this step was repeated twice. The elutes were snap frozen in liquid N_2 and vacuum dried. Samples were then dissolved in 3% acetonitrile and 0.1% formic acid, briefly sonicated in a sonic water bath for 5mins, and incubated for 10mins at RT.

Samples were further processed by Joel Wirz (University of Zurich) at Functional Genomics Centre Zurich, as follows. Dissolved samples were injected by a Waters M-class UPLC system (Waters AG) and separated on a C18 reverse-phase column. High accuracy mass spectra were acquired with a Q-Exactive HF mass spectrometer (Thermo Scientific) that was operated in data-dependent acquisition mode. A survey scan was followed by up to 12 MS2 scans.

2.15 Data analysis

2.15.1 Statistical analyses

All measurements of animal samples were biological replicates and represent individual animals. Cell culture replicates are technical replicates, unless indicated that samples represent individually derived clones. Basic statistical tests used are described in figure legends, where applicable.

Non-parametric Kruskall-Wallis ranked sum test followed by Wilcoxon sum rank test were used for datasets where mean values of genomic tiles are compared across more than two genotypes. Notably, these analyses are heavily overpowered due to exceptionally large number of variables and almost always show strong significance. The test results are listed in the figure legends where applicable for thoroughness of data analysis. However, when mean values for two or more genotypes are not showing more than net 10% methylation differences, they are considered not biologically significant. For detection of differentially expressed/methylated genes/tiles statistical tools designed specifically for the type of data were used.

Analysis of high-throughput sequencing samples was done using Seqmonk visualisation and analysis tool, using default settings unless otherwise indicated.

2.15.2 RNA-seq analysis

RNA-seq data were trimmed using TrimGalore 0.5.083 or newer versions, then mapped using HiSat2 2.1.084 or newer versions to Mus musculus genome GRCm38. Processed data was analysed using Seqmonk 1.42.085 or newer versions. Reads over transcripts were merged across exons correcting for feature length, quantitated using RNA-Seq quantitation pipeline and log_2-transformed. Differentially expressed genes were called using DESeq with significance threshold of $p<0.01$ after Benjamini-Hochberg multiple testing correction. Mitochondrial genes were excluded. For genome browser views of gene expression, selected regions were split into 500bp bins every 50bp and quantitated as reads per million transcripts, unless specified otherwise. David 6.786 Functional Annotation Tool was used for gene ontology analysis, with categories GOTERM_BP_FAT, GOTERM_CC_FAT, GOTERM_MF_FAT, using Benjamini-corrected $p<0.01$ significance cut-off.

2.15.3 Methylation data analysis

Methylation data were trimmed using TrimGalore 0.5.083, deduplicated, mapped to Mus musculus genome GRCm38, and methylation calls were extracted using Bismark 0.19.187 or newer versions. Processed data were analysed using Seqmonk 1.42.085 or newer versions. Pituitary datasets were down-sampled to the least covered dataset using SeqMonk to avoid differential coverage artefacts. Probes were defined as tiles of CpG positions and quantified using Bisulphite quantitation pipeline. Tiles in the oocytes, adult hypothalamus and neuronal progenitor cells were 100-CpGs where at least 10 are covered. In pituitary, liver, E7.5 epiblast and when E7.5 epiblast data was used in analysis, tiles of 300-CpGs with at least 10 CpGs covered were used. X and Y genes were excluded only from E7.5 epiblast analyses. Differentially methylated regions were called using Seqmonk edgeR (for/rev) statistics filter with absolute methylation difference cut-off of 20%, using multiple-testing corrected p-value ≤ 0.01. Genomic locations were defined as listed in the brackets using feature annotations available in SeqMonk: promoter (-1500bp to +500bp around transcription start site), gene body (+500bp to the end of the mRNA/gene), intergenic (promoters and gene bodies excluded). CpG island features were called based on previously published CpG island coordinate data (Illingworth et al., 2010). For promoter analysis, 100CpG windows were fused together if distance between them were less than 1kb. Random probes were selected from all probes, excluding those in the DMR list, and utilised as representative of the genome. Previously described DNA methylation 'canyon' coordinates were used in DNA methylation valley analysis (Jeong et al., 2014) and only those showing methylation under 10% and covered by at least 5 CpGs in hypothalamus were included. Non-CpG methylation was analysed using 1000CpH tiles where at least 30 CpHs were covered.

2.15.4 Ultra-low input and biotin ChIP data analysis

ChIP-seq data were trimmed using Trim Galore 0.5.083 and mapped against *Mus musculus* GRCm38 genome with Bowtie 1.2.290. H3K27me3 enrichment was quantitated using 2kb tiles with a 1kb step, correcting for total count based on the largest library, unless stated otherwise. H3K4me3 and H3K27me3 peaks were called using MACS peak caller within SeqMonk 1.42.0 or newer versions, using a significant threshold of $p<0.00001$, a sonication fragment size of 1kb, and input controls as a reference. When assessment of histone marks was done in relation to DNA methylation, DNA methylation tiles were used to quantitate the enrichment.

2.15.5 Label-free mass spectrometry data analysis

Protein identification from MS-raw data was performed using MaxQuant 1.6.12.0 and analysed using *Proteus* package in R statistical software. Peptide aggregation was done by Leading Razor Protein method (Silva et al., 2006), and samples were normalised by matching median intensities. Pairwise statistical comparison was done using limmaDE statistical tool.

3 DNA methylation in oocytes carrying the *Dnmt3a*-D329A mutation

Declaration of author contributions

Oocyte collections were done together with Courtney Hanna.

Mammalian oocytes show a unique DNA methylation pattern, where only actively transcribed gene bodies are highly methylated. The rest of the genome, unlike in a somatic cell, is largely unmethylated, thus forming hyper- and hypo- DNA methylation domains (Smallwood et al., 2011; Kobayashi et al., 2012; Veselovska et al., 2015). Moreover, only actively transcribed genes in the oocyte are marked by histone 3 lysine 36 trimethylation (H3K36me3), which correlates highly with DNA methylation (Xu et al., 2019). Despite the unique methylome, there is no specialised DNA methylation machinery in the mouse oocyte (Kaneda et al., 2010). A shorter isoform DNMT3A2 together with a germ-cell specific co-factor DNMT3L is responsible for DNA methylation establishment, while DNMT3B does not play a significant functional role. PGCs undergo epigenetic landscape erasure and the unusual oocyte-specific methylation pattern is solely a result of *de novo* methylation events occurring during a meiotic arrest. These circumstances create a perfect model to study the impact of the DNMT3A D329A mutation on DNA methylation targeting to the H3K36me3-marked fraction of the genome.

3.1 Assessment of global DNA methylation in oocytes

In order to assess whether DNA methylation in the oocyte is affected in the presence of the D329A mutation, germinal vesicle (GV) oocytes were collected from 24-26 day old *Dnmt3a*$^{+/+}$, *Dnmt3a*$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$, and *Dnmt3a*$^{\Delta/D329A}$ female mice coming from one litter, as described in the breeding scheme in section 1.12. Briefly, females used in the cross were heterozygous for a floxed *Dnmt3a*^{fl} allele and carried oocyte-specific Zp3

promoter driven *Cre* recombinase (Kaneda et al., 2010), which results in offspring inheriting either the *wild-type* allele or the deletion (Δ) of *Dnmt3a* (Figure 1.6). Due to challenges of maternal passage of mutant allele, it was not possible to create a homozygous mutant, therefore the mutant allele was passed from heterozygous *Dnmt3a*$^{+/D329A}$ males. I used all four genotypes to account for potential haploinsufficiency of *Dnmt3a*, therefore, if needed, allele number matched pairs could be compared.

Approximately 75 to 170 oocytes were collected from an individual mouse and these would constitute one sample, which was processed using post-bisulphite adaptor tagging (PBAT) to generate oocyte methylome maps. In total three to four litter-matched mice were used per genotype. Libraries had relatively low levels of duplication, ranging from 12.5% to 20.4% and had between 19 and 39 million unique deduplicated alignments (Table 1). They all passed FastQC checks of raw sequencing data.

To evaluate the PBAT datasets, the genome was split into tiles containing 100 CpG dinucleotides, and only tiles with methylation calls at least 10 CpG dinucleotides were used in the analysis. Initially samples were assessed by drawing a similarity tree and a correlation matrix, which are based on Pearson distance and pairwise correlations, respectively (Figure 3.1). Sample *+/D329A_2* showed the lowest correlation with other samples, as it branched off from the main distance tree immediately (Figure 3.1a). It also had uniquely low pairwise correlation compared to other sample pairs (Figure 3.1b). Sample *+/D329A_2* also showed a spatial separation in two-dimensional space in Principal Component Analysis (PCA) (Figure 3.2a). Given the evidence of dissimilarity to all other samples in this experiment sample *+/D329A_2* was excluded from further investigation. This has improved overall spatial separation of the rest of the samples on PC2 (Figure 3.2b).

Table 1. Oocyte PBAT library summary.

Sample ID	Duplication rate	Unique deduplicated alignments
+/+_1	12.5%	38,386,803
+/+_2	15.1%	22,558,552
+/+_3	14.9%	19,325,117
+/+_4	15.5%	23,832,813
Δ/+_1	17.9%	28,438,378
Δ/+_2	17.7%	32,110,878
Δ/+_3	16.1%	23,390,327
Δ/+_4	14.5%	39,466,786
+/D329A_1	15.6%	33,928,176
+/D329A _2	16.9%	28,766,719
+/D329A _3	20.4%	29,092,007
Δ/D329A _1	15.1%	33,621,919
Δ/D329A _2	16.9%	25,347,250
Δ/D329A _3	13.5%	24,815,042
Δ/D329A _4	14.5%	28,320,511

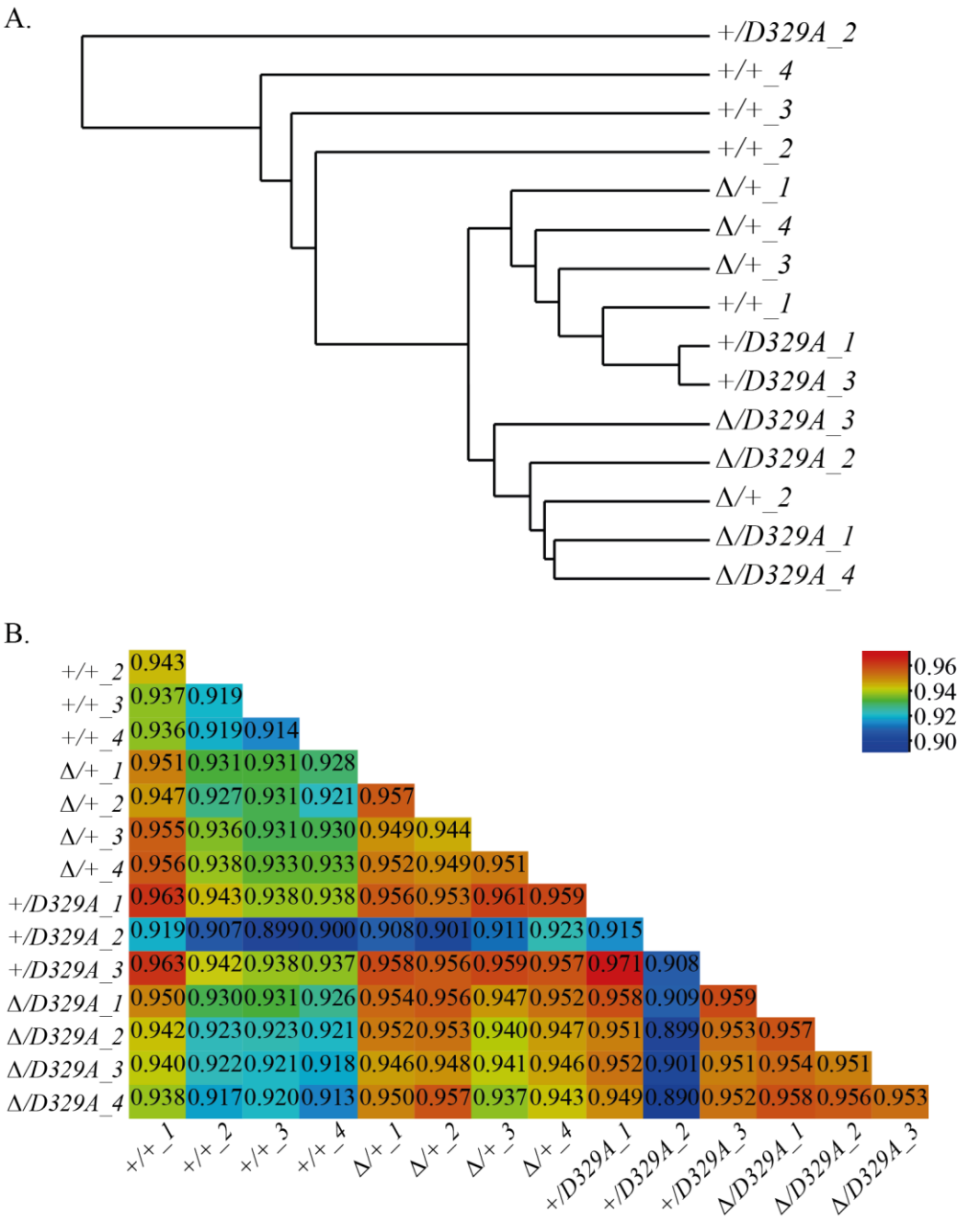


Figure 3.1. Bulk oocyte sample correlation. A. A data similarity tree indicating Pearson correlation distances between individual samples of the genotypes listed. **B.** Pairwise correlation matrix for individual sample pairs, in scale 0 to 1. Numbers denote replicates.

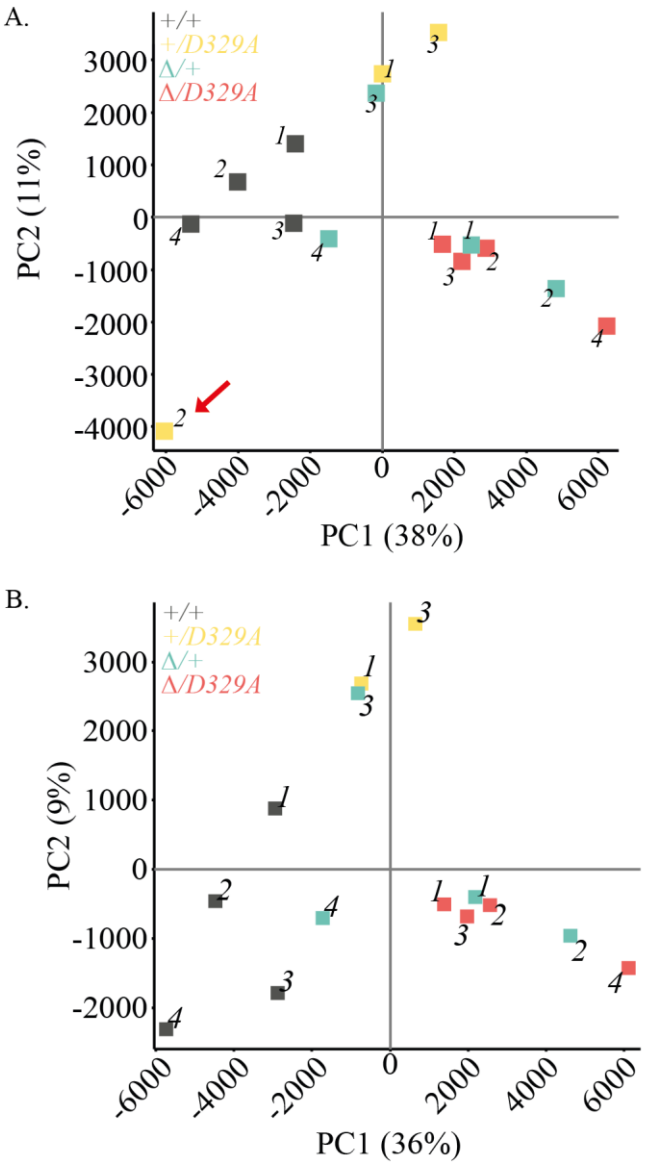


Figure 3.2. Principal Component Analysis of oocyte samples. PCA was done on 100CpG tiles, where at least 10CpGs are covered. Each dot represents a replicate sample coming from an individual mouse, colour denotes the genotype. PCA plot in **A.** is before exclusion of *+/D329A_2* sample and **B.** is after.

Global DNA methylation in all samples was similar, abeit not identical, and showed a distribution usual for the oocyte, where most of the tiles were either hyper- or hypo-methylated (Figure 3.3). Mean methylation was 42.9%, 36%, 37.9% and 34% in *Dnmt3a*$^{+/+}$, *Dnmt3a*$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$, respectively. Global levels of approx. 7% lower DNA methylation in *Dnmt3a*$^{\Delta/+}$ compared to *Dnmt3a*$^{+/+}$ oocytes indicated the effect of haploinsufficiency. In agreement with this, *Dnmt3a*$^{\Delta/D329A}$ samples

had approx. 4% lower DNA methylation than *Dnmt3a*$^{+/D329A}$, suggesting that dosage of the *wild-type* allele was important.

When genomic tiles of the *Dnmt3a*$^{+/+}$ samples were aligned against *Dnmt3a*$^{\Delta/D329A}$ a shift in the correlation axis on the scatterplot could be observed, suggesting that many genomic locations are more methylated in the *wild-type* sample (Figure 3.4a). A similar trend was observed when comparing *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/+}$ (Figure 3.4b). A comparison between *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$ or *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$ did not show this curved effect of the correlation axis, suggesting that a tendency for hypomethylation observed was caused by haploinsufficiency rather than presence of the D329A mutation (Figure 3.4).

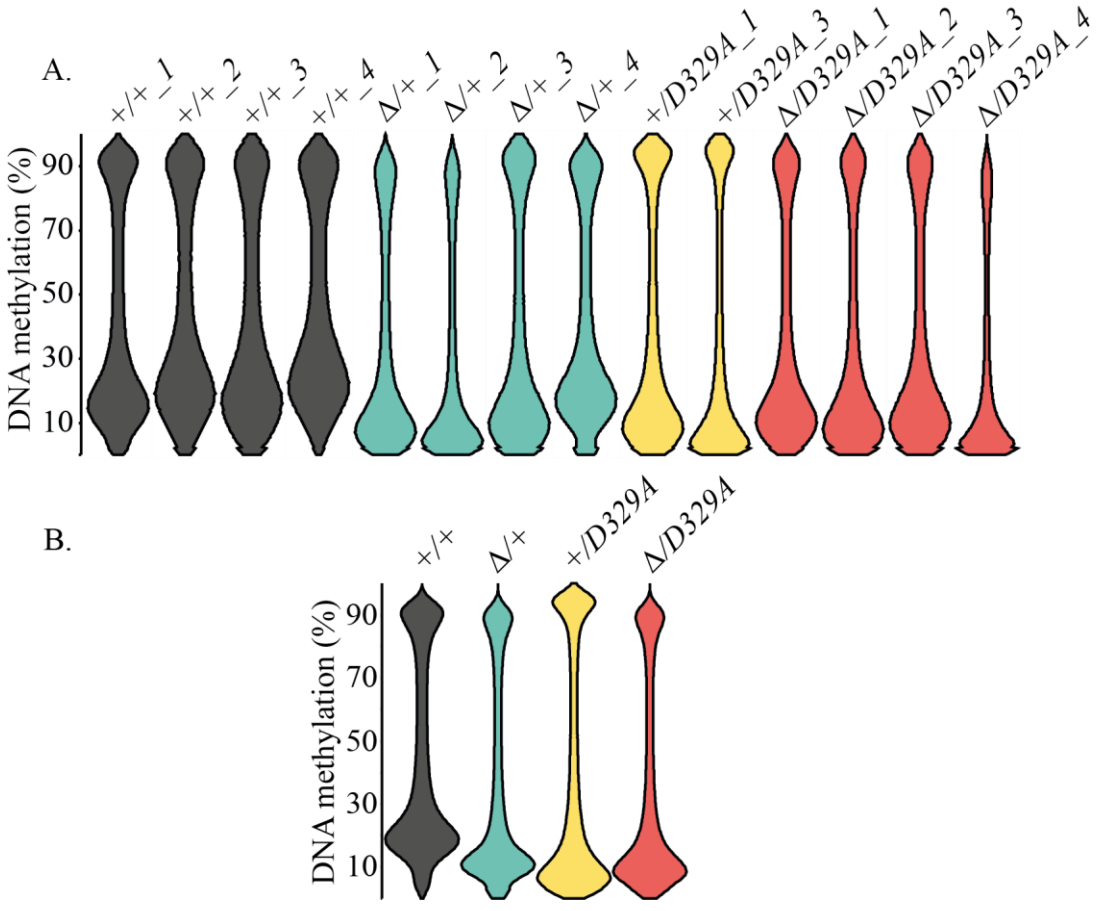


Figure 3.3. Global oocyte DNA methylation. A. DNA methylation levels in individual samples. **B.** DNA methylation levels averaged within the genotype. 100CpG tiles with coverage in all samples were used (total of 395,014 tiles). Kruskal-Wallis rank rum test resulted in p-value < 2.2e-16, and was significant for all post-hoc pairwise Wilcoxon rank sum tests.

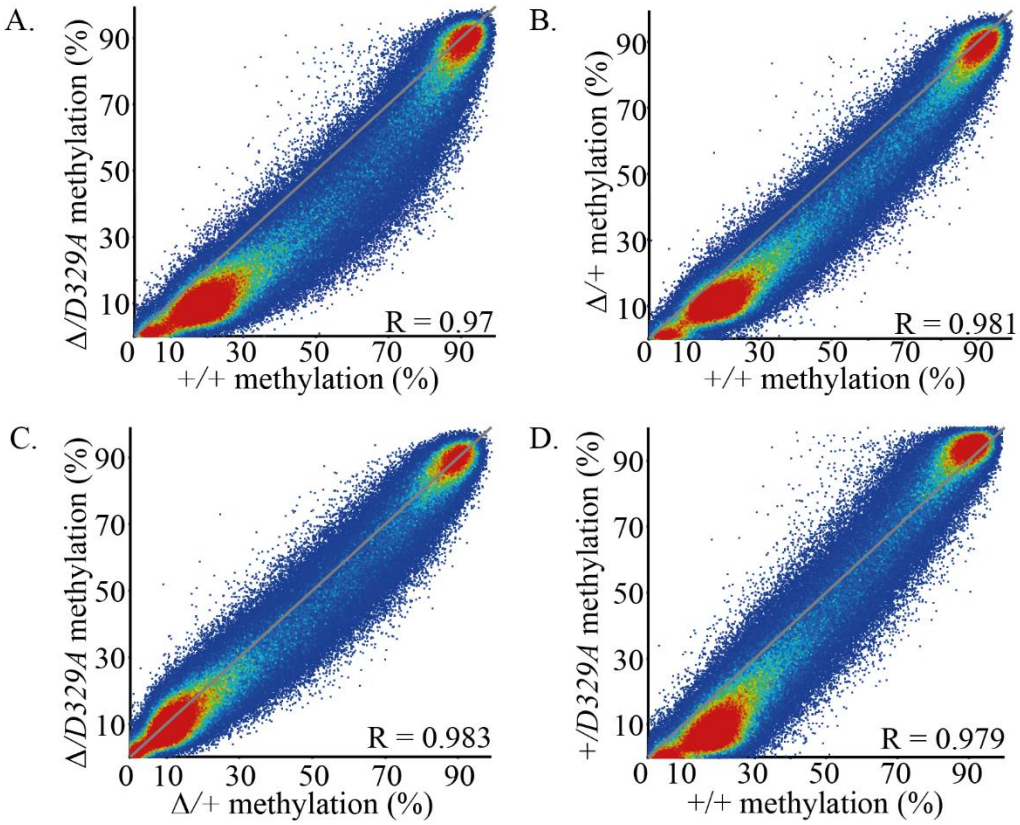


Figure 3.4. Scatterplots showing alignment of genomic regions between the different genotypes. Genotypes are listed on *x* and *y* axes. Each dot represents a genomic tile covering 100 CpGs, mean methylation value is merged across the replicates.

3.2 Investigating possible somatic cell contamination

It is evident in Figure 3.3 that the genotype replicates did not have identical distributions of methylation of 100-CpG tiles. I considered the possibility that somatic cell DNA contamination could contribute to this variation. Oocyte collection is conducted manually by dissociating ovaries using tissue digestion enzymes and pipetting force and by passing a piece of tissue through gradually smaller diameter pipette tips. The lysis and shear force can burst cells and lead to release of some cell nucleus material into the collection buffer, resulting in free-floating DNA. Additionally, oocytes are surrounded by granulosa cells, many times smaller and substantially more difficult to see under a benchtop microscope. As there were on average 100 oocytes in one sample, some contamination by soma is difficult to avoid. Since the X chromosome is silenced by DNA methylation in somatic cells but not in the oocyte, DNA methylation of X chromosome CGIs can be used as a proxy for contamination assessment.

Assessment of genomic tiles overlapping X chromosome CGIs showed that some of the samples had higher levels of DNA methylation, suggesting presence of some somatic cell DNA contamination (Figure 3.5, Table 2). This was particularly pronounced in *wild-type* oocytes, showing >20% mean DNA methylation of X chromosome CGIs, and especially in sample *Dnmt3a*$^{+/+}$*_4*, where X chromosome CGIs are methylated to 27.8%.

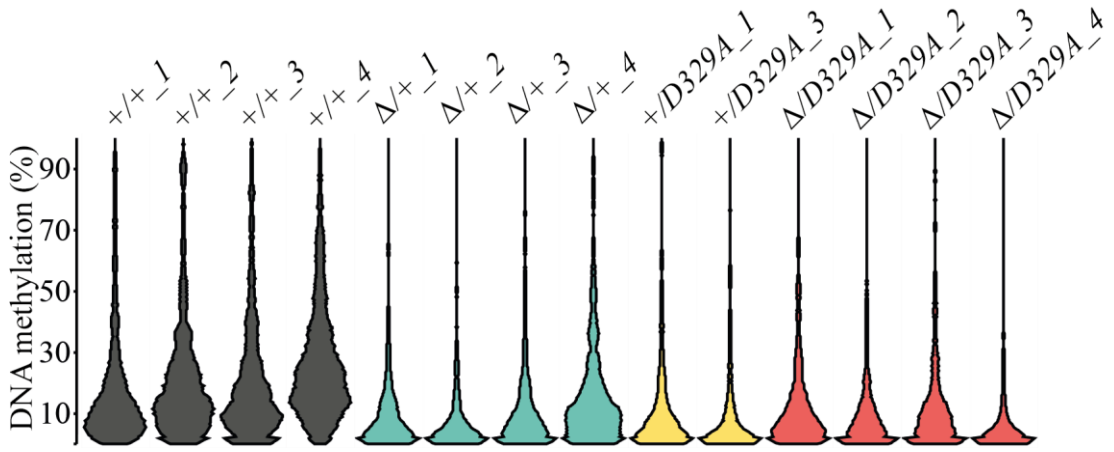


Figure 3.5. DNA methylation of tiles overlapping X chromosome CGIs. Individual replicates are shown. Tiles of 100CpG overlapping CGIs were used.

Table 2. DNA methylation levels of X-chromosome CGIs. Individual samples are listed. In total 971 X chromosome CGIs with coverage in all samples were assessed.

Sample ID	**X-chr CGI DNA methylation**
Dnmt3a$^{+/+}$_1	20.4%
Dnmt3a$^{+/+}$_2	22.9%
Dnmt3a$^{+/+}$_3	22.2%
Dnmt3a$^{+/+}$_4	27.8%
Dnmt3a$^{\Delta/+}$_1	14.0%
Dnmt3a$^{\Delta/+}$_2	12.7%
Dnmt3a$^{\Delta/+}$_3	15.7%
Dnmt3a$^{\Delta/+}$_4	18.28%
Dnmt3a$^{+/D329A}$_1	16.2%
Dnmt3a$^{+/D329A}$_2	14.8%
Dnmt3a$^{\Delta/D329A}$_1	16.0%
Dnmt3a$^{\Delta/D329A}$_2	14.1%
Dnmt3a$^{\Delta/D329A}$_3	15.8%
Dnmt3a$^{\Delta/D329A}$_4	10.8%

Initially, it was hypothesised that, due to the introduction of the D329A mutation, oocytes would lose a relatively large proportion of DNA methylation specifically over gene bodies. Given that somatic contamination would likely result in evenly distributed elevation of DNA methylation across the whole genome, it is unlikely that small levels of contamination would mask this effect completely. A visual assessment of genome browser tracks for groups gave confidence that minor contamination would not hinder further analysis (Figure 3.6).

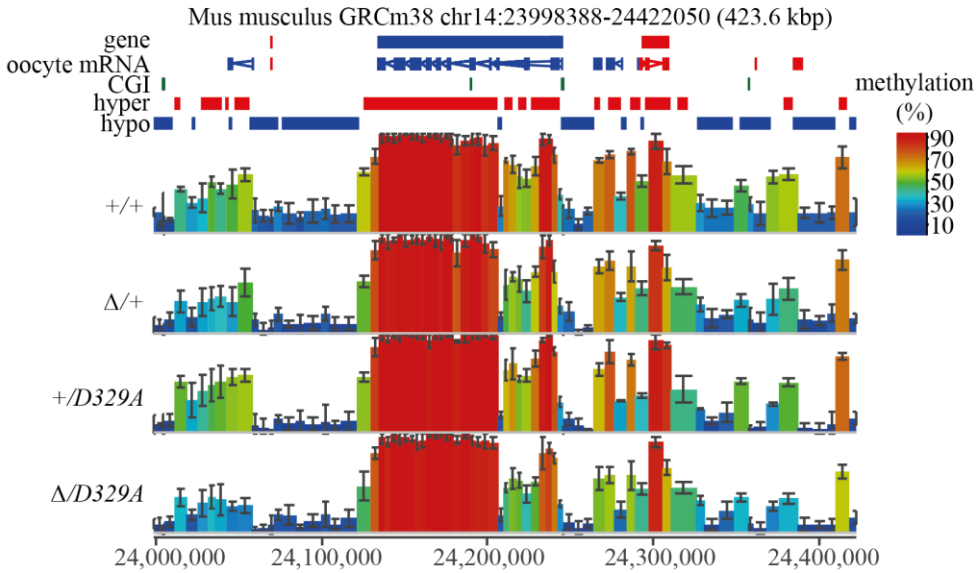


Figure 3.6. Genome browser view of DNA methylation in oocytes. Each tile represents 100-CpG average methylation values; error bars indicate error of the mean. Hyper refers to hypermethylated domains in the oocyte, hypo – domains showing low methylation levels.

A prominent feature of the oocyte methylome is the presence of hypo- and hyper-methylated domains, and this pattern was clearly seen in all four genotypes in the genome browser view (Figure 3.6). Assessment of DNA methylation levels over these domains showed minor differences across the phenotypes (Figure 3.7). Hypermethylated domain DNA methylation levels appeared to be dependent on the allele number, on average samples with two alleles had 9% higher DNA methylation levels than samples with one allele only (Table 3). Meanwhile, the difference between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$,

or *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$ samples was in the range of 1.7-4.2%. Hypomethylated domains had DNA methylation within the range of 12.2-14.6%, with an exception of the *Dnmt3a*$^{+/+}$ genotype, which was methylated to 22.6% at hypo- domains. This could be a result of higher contamination levels in *wild-type* samples. Overall, the variation across genotypes, especially over hypermethylated oocyte domains, did not indicate major global shifts.

Table 3. Mean DNA methylation levels of hyper- and hypo- methylated domains. Numbers represent group average.

Genotype	**Hypermethylated domain**	**Hypomethylated domain**
Dnmt3a$^{+/+}$	80.0%	22.6%
Dnmt3a$^{\Delta/+}$	74.0%	14.6%
Dnmt3a$^{+/D329A}$	81.7%	13.6%
Dnmt3a$^{\Delta/D329A}$	69.8%	12.2%

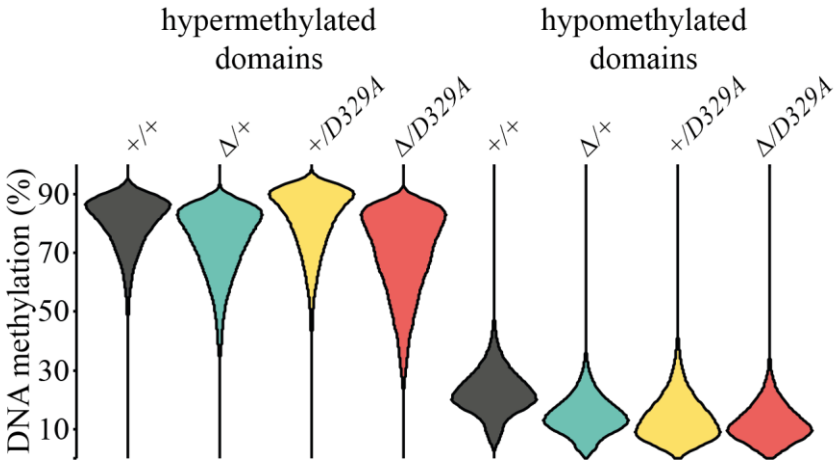


Figure 3.7. DNA methylation levels of hyper- and hypo- methylated domains. Tiles of 100CpGs with coverage in all samples were used, replicates are grouped. For both hypermethylated and hypomethylated domains Kruskal-Wallis rank rum test p-value was < 2.2e-16, and significant for all post-hoc pairwise Wilcoxon rank sum tests.

In addition to hypo- and hyper- methylated domains I assessed other genomic features. DNA methylation levels over tiles overlapping CGIs or CGI-containing promoters were very similar (Figure 3.8a,b; Table 4). Meanwhile non-CGI promoters had 23.7-28% higher DNA methylation level than CGI promoters (Figure 3.8b). Overall, the variation in DNA methylation levels across all genomic features corresponded to the number of *Dnmt3a* alleles. Gene bodies were generally highly methylated across all four genotypes; again only a minor decrease was observed when one copy of *Dnmt3a* was missing. Meanwhile, intergenic domains had very low methylation, as expected of the oocyte methylome (Figure 3.8c). No dramatic differences in the presence of the D329A mutation could be seen. Notably, the *Dnmt3a*$^{+/+}$ oocytes had the highest DNA methylation in all comparisons depicted in Figure 3.8. This, and the lower level of DNA methylation observed previously in hyper- domains, could be explained as an artefact of higher somatic cell DNA contamination levels found in some of those samples.

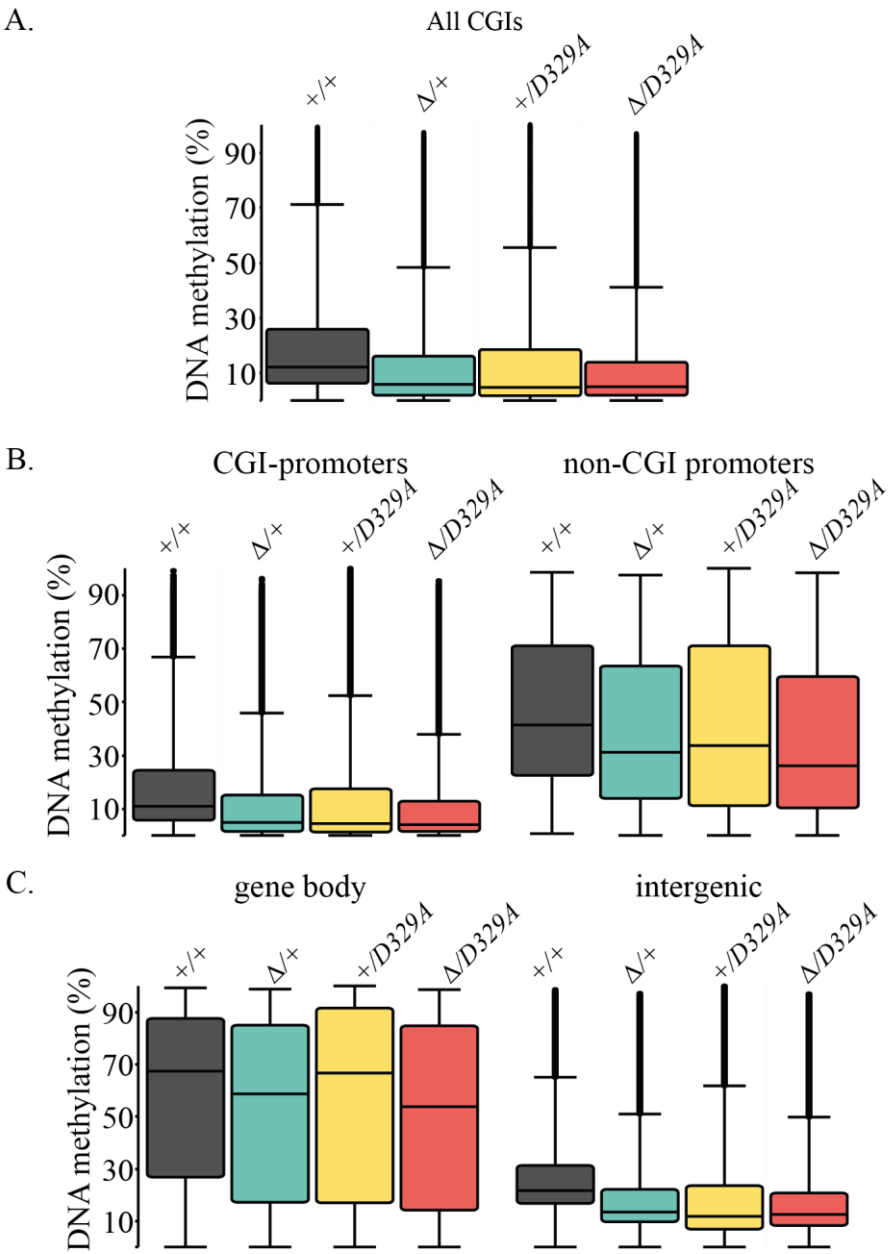


Figure 3.8. DNA methylation levels of various genomic features. DNA methylation levels of **A.** all CGIs, **B.** CGI-overlapping and non-CGI promoters, **C.** gene bodies and intergenic regions. 100 CpG tiles overlapping features indicated were used in quantification. Kruskal-Wallis rank rum test p-value was < 2.2e-16, and significant for all post-hoc pairwise Wilcoxon rank sum tests.

Table 4. DNA methylation levels over oocyte genomic features. Mean for each genotype group is provided.

Genotype	All CGIs	CGI promoters	Non-CGI promoters	Gene body	Intergenic
Dnmt3a$^{+/+}$	21.3%	18.9%	46.9%	58.6%	27.4%
Dnmt3a$^{\Delta/+}$	15.3%	13.0%	39.4%	52.5%	19.7%
Dnmt3a$^{+/D329A}$	16.8%	14.5%	41.2%	56.5%	19.4%
Dnmt3a$^{\Delta/D329A}$	14.6%	12.2%	35.9%	50.2%	17.9%

3.3 Link between oocyte DNA methylation and histone PTMs

Given that hypermethylated domains in the oocyte correlate with H3K36me3 (Xu et al., 2019), I next assessed DNA methylation over different histone modifications. Previously published datasets for H3K4me3 and H3K27me3 (Hanna et al., 2018) and an unpublished H3K36me3 dataset, courtesy of C. Hanna, were used as a reference for high (top 10%) and low (bottom 10%) enrichment genomic areas in the oocyte.

First, I looked at areas that are enriched for H3K36me3. High levels of H3K36me3 correlated with DNA methylation; on average, these genomic regions had 77.5% DNA methylation, while areas with low H3K36me3 enrichment were methylated only to 23.6% (Figure 3.9, Table 5). Given that the difference across any two genotypes over low- or high- H3K36me3 enriched domains did not exceed 9%, data again confirmed the lack of predicted loss-of-function effect of D329A mutation in relation to H3K36me3.

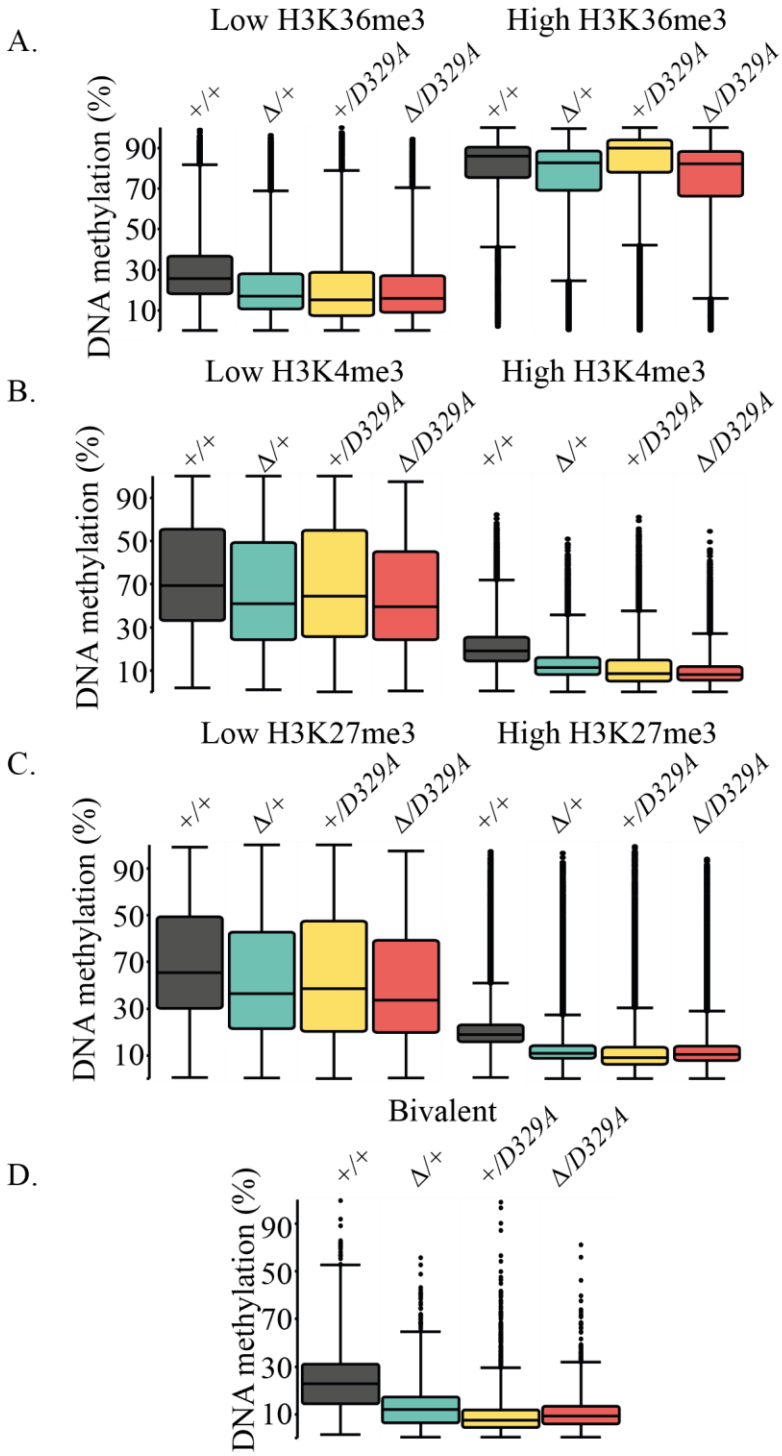


Figure 3.9. DNA methylation levels at genomic areas linked to histone PTMs. DNA methylation distribution in genomic tiles linked to **A.** H3K36me3, **B.** H3K4me3, **C.** H3K27me3, **D.** bivalent chromatin (high H3K4me3 and high H3K27me3). High H3K4me3 and high H3K27me3 datasets exclude bivalent tiles. Low refers to bottom 10 % and high to top 10% of enrichment of genome-wide PTM distribution. Kruskal-Wallis rank rum test p-value was < 2.2e-16, and significant for all post-hoc pairwise Wilcoxon rank sum tests. The exception was **C.** low H3K27me3 areas, where post-hoc Wilcoxon

rank sum pairwise test for *Dnmt3a*$^{\Delta/+}$ against *Dnmt3a*$^{+/D329A}$ p-value (Benjamin Hochberg adj.) = 0.079.

Generally, similar trends of DNA methylation were observed in genomic areas with low- or high enrichment for H3K4me3, H3K27me3. Broadly, tiles highly enriched for H3K4me3 had approximately 14.0% DNA methylation, while areas with low H3K4me3 enrichment were on average 48.8% methylated. Similarly, high H3K27me3 enriched areas had approx. 15% methylation, while low H3K27me3 enriched areas had on average 44.3% methylation. Bivalent chromatin was on average 8.5% methylated. In all datasets, *wild-type* samples were the ones with highest DNA methylation and the main driver of intergenotype-based variability. Overall, these data suggest that the D329A mutation does not lead to differences in DNA methylation levels over genomic tiles enriched for specific histone PTMs.

Table 5. Mean levels of DNA methylation in the oocyte genomic areas linked to histone PTMs. Genotype column indicates alleles of *Dnmt3a* gene. High H3K4me3 and high H3K27me3 datasets exclude bivalent tiles. Bivalent contains those genomic tiles that show high H3K4me3 and high H3K27me3. Low refers to bottom 10 % and high to top 10% of enrichment distribution.

***Dnmt3a* Genotype**	**Low K36 me3**	**High K36 me3**	**Low K4 me3**	**High K4 me3**	**Low K27 me3**	**High K27 me3**	**Bivalent**
+/+	29.9%	79.5%	53.7%	21.1%	49.9%	21.3%	14.5%
Δ/+	22.3%	75.1%	46.8%	13.2%	42.8%	13.4%	7.6%
+/D329A	21.4%	81.7%	49.7%	11.7%	44.3%	12.6%	5.5%
Δ/D329A	20.9%	73.8%	45.2%	9.9%	40.3%	12.6%	6.2%

3.4 Assessment of single-cell oocyte samples

Bulk oocyte analysis revealed little to no effect of the presence of DNMT3A^{D329A} on DNA methylation. However, the interpretation was likely to be confounded by varying levels of somatic cell or cell-free contamination. In order to get a more conclusive answer, GV oocytes were collected for single-cell DNA methylation analysis. Somatic cells are methylated to approx. 70%, while oocyte global DNA methylation is around 40%. Even if there was a single somatic cell present, it would immediately skew the global DNA methylation level in a sample and therefore it could be easily identified and removed from analysis. In addition, oocytes of 20-week old females were collected to account for a possibility of progressive accumulation of aberrant DNA methylation during aging. Finally, matched *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$ littermates were used to avoid the confounding dosage effect which was observed in bulk analysis.

Approximately 20 GV oocytes were collected from two mice of each *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$ genotype and processed by single-cell bisulphite-sequencing (scBS-seq). Samples showing no alignment to *Mus musculus* genome were excluded, resulting in 21 *Dnmt3a*$^{+/+}$ and 19 *Dnmt3a*$^{+/D329A}$ oocytes. Full genome coverage is not possible in single-cell bisulphite sequencing due to DNA degradation, PCR biases, allele drop-outs and sequencing costs; it is only possible to recover information on limited CpG sites (Clark et al., 2017). Single oocyte libraries had approx. 1 to 4.4 million reads, and approx. 2.41% to 10.85% of all CpG sites were covered (individual numbers provided in Appendix 1).

Given that many cells are assessed and genomic regions covered in each one of them would be random, I first looked at DNA methylation quantification of individual cells, regardless of whether the genomic tiles assessed were covered in the majority samples or not. Pearson correlation was used to assess similarity of individual samples. First, a relationship tree based on correlation distance showed one sample – 7_+/D329A_310.1e_1 – as an outlier (Figure 3.10). Pairwise correlation values between most samples were ≥0.65, while the previously indicated outlier showed ≤0.51 correlation to other samples (data not shown). Based on this evidence, sample 7_+/D329A_310.1e_1 was excluded from further analysis. Notably, there was no clear separation of the samples based on the processing time (so called batch-effect) or coverage of individual samples (Figure 3.11), which was reassuring that the samples are of good quality. Global DNA methylation values, independently of genotype, ranged from 27.4% to 36.2%, while X chromosome CpG island DNA methylation was between 2.7% to 8.5% (Figure 3.11).

Together, this suggested there was no somatic cell DNA contamination, no batch or sequencing depth effects in single oocyte samples.

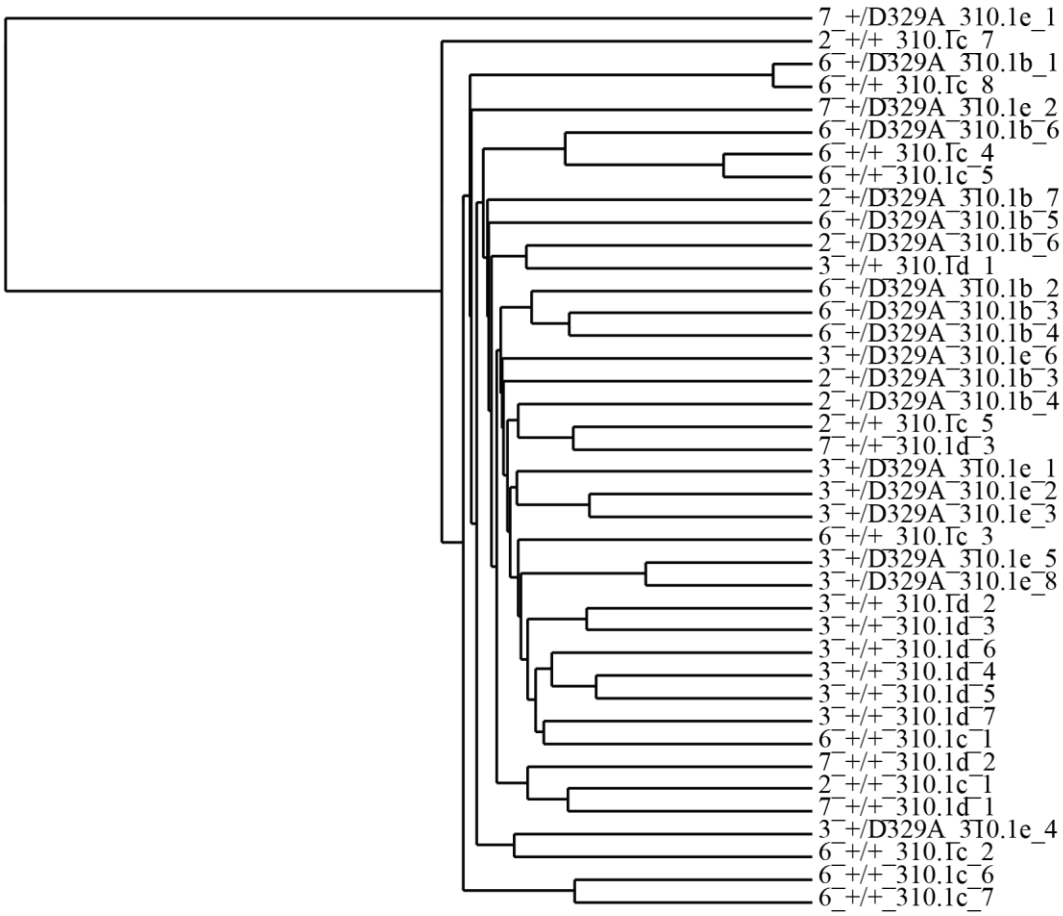


Figure 3.10. Similarity tree of individual oocyte samples. The tree is drawn based on Pearson distance correlation.

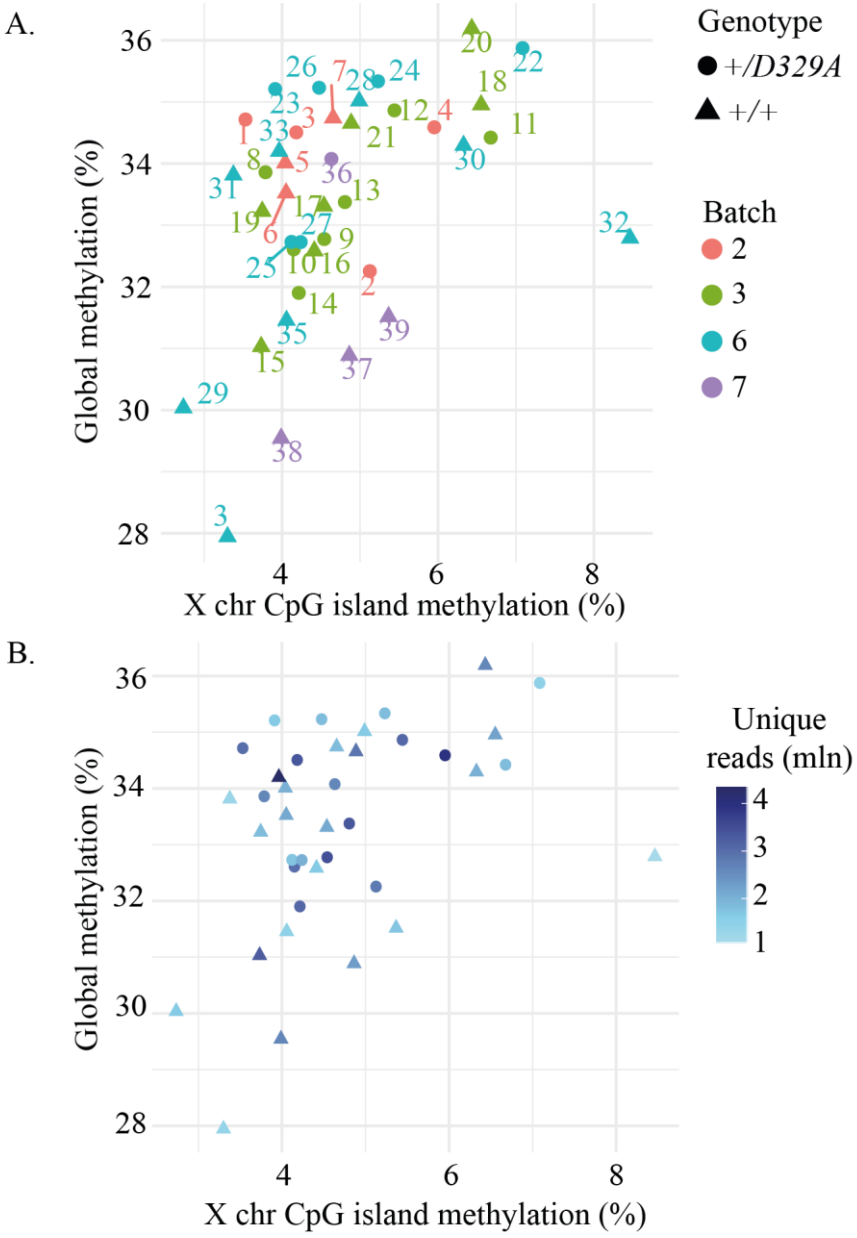


Figure 3.11. Single cell oocyte DNA methylation correlation to X-chr CGIs. Global DNA methylation relationship to X chromosome CGIs is visualised as a scatterplot. Shape denotes genotype, as described in the legend. **A.** Colour denotes cell collection batch. **B.** Colour denotes number of unique read millions included in the analysis for individual samples. Numbers refer to single oocyte samples, corresponding sample names are provided in Appendices.

Given that oocytes are a relatively homogeneous cell population and inter-genotype variability is not expected, I could benefit from using confidently clean single oocyte datasets combined together into groups resembling bulk oocyte samples. I elected to

group together 4-6 samples into four pseudo-bulk groups per genotype in order to obtain more informative datasets. Individual samples were assigned randomly, to ensure the final counts of mapped reads in all four pseudo-bulk groups were as similar as possible, between 10.6-11.8 $x10^6$ (list is provided in Appendices). The coverage of CpG positions in pseudo-bulk samples, compared to individual ones, has increased from 2.7%-8.5% to 23.2%-25.4%, therefore providing more even and extensive coverage of the genome. Since the genome was separated into sequential tiles containing 100 CpG positions, and those with at least 10 individual CpG positions covered were used in quantification, usage of pseudo-bulking approach resulted in 160,437 out of total 233,658 genome tiles being covered in all eight pseudo-bulk samples. These were used in subsequent analysis, therefore making the samples directly comparable.

Independently of the genotype, Pearson correlation between the groups was ≥0.93, while assessment of the PCA plot indicated a spatial separation of samples based on the genotype (Figure 3.12). Global DNA methylation exhibited the expected bimodal distribution, and there were no differences between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$ pseudo-bulk samples, nor was there any shift in correlation axis when genomic tiles with average genotype values were aligned (Figure 3.13). Overall, the pattern of DNA methylation was clear and robustly similar in all samples, with no evidence of somatic DNA contamination over intergenic domains (Figure 3.14). Oocyte hypo- and hyper-methylated domains also showed regular distribution in both genotypes, with no visible variation across the pseudo-bulk samples (Figure 3.15).

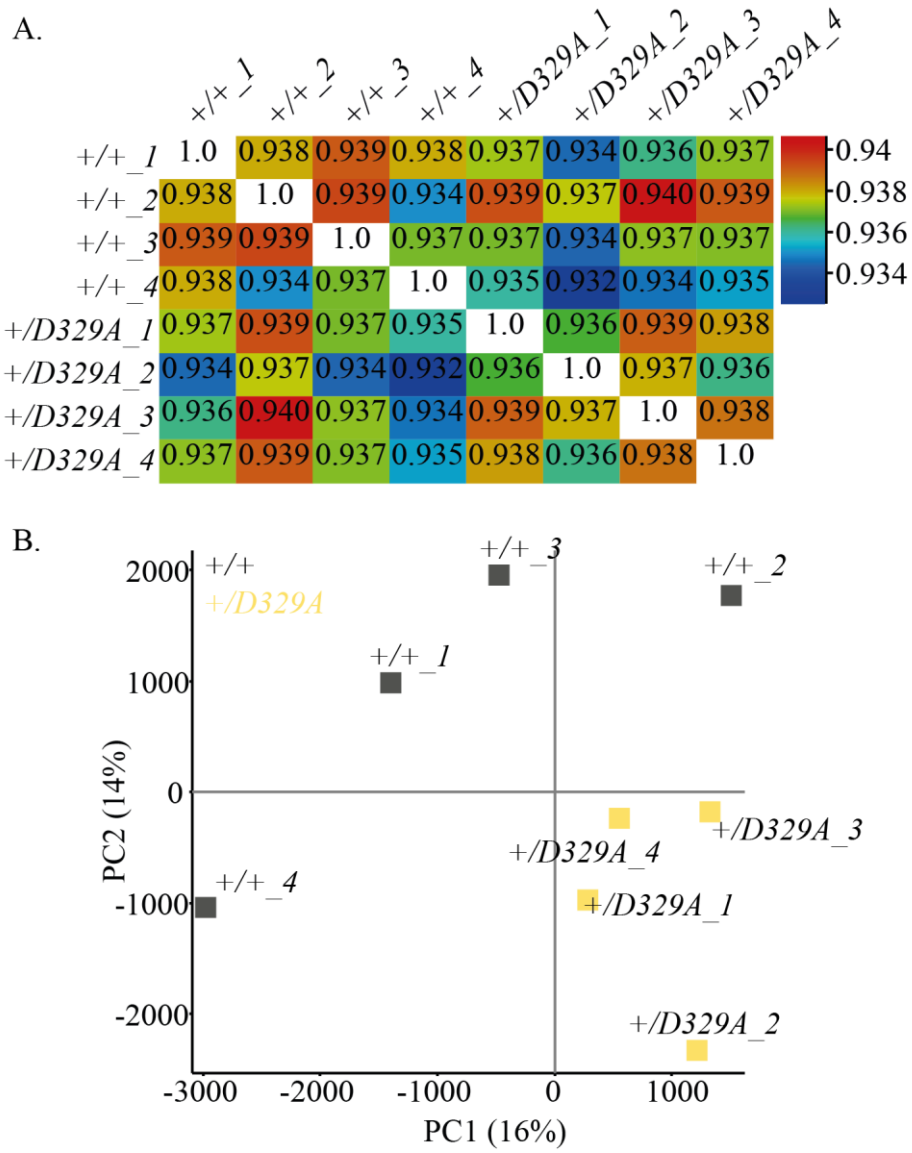


Figure 3.12. Correlation of pseudo-bulk oocyte samples. A. Pearson pairwise correlation of pseudo-bulk oocyte samples. **B.** PCA plot showing separation of individual pseudo-bulk samples. Colour denotes genotype, *Dnmt3a* alleles are provided as sample names.

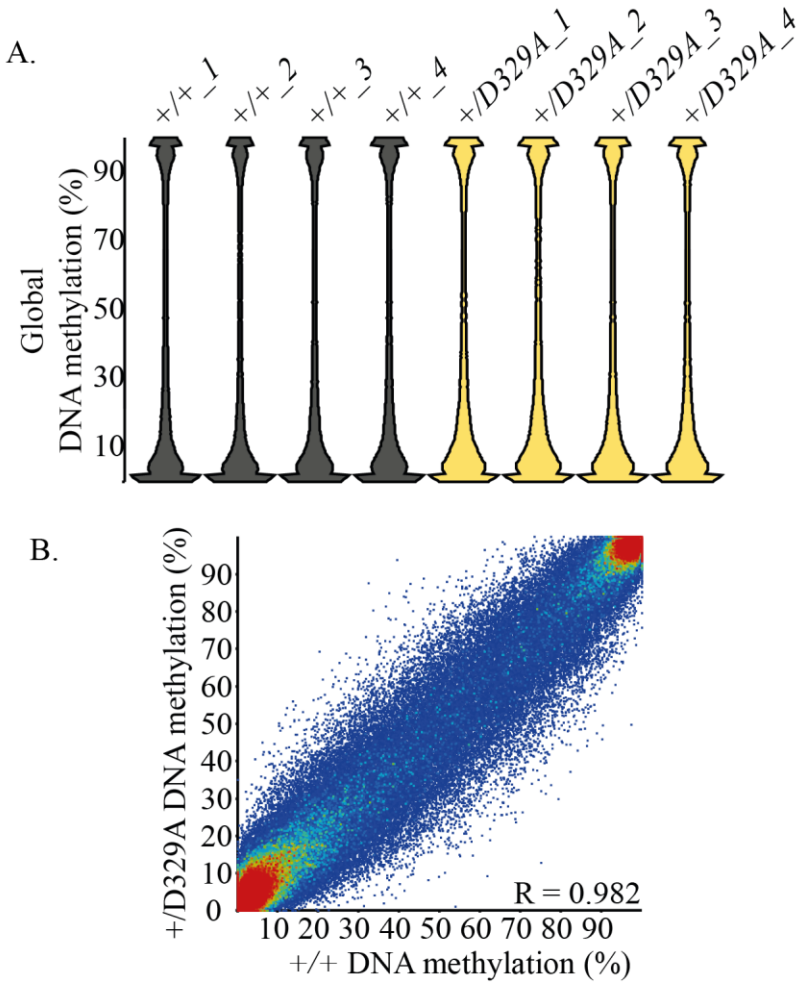


Figure 3.13. Global DNA methylation of pseudo-bulk oocytes. Alleles of *Dnmt3a* are listed as sample names. Genomic tiles of 100CpGs with at least 10 positions covered were used. **A.** Beanplots showing global DNA methylation distribution of individual pseudo-bulk samples. Wilcoxon rank sum test using mean tile values for each genotype results in p-value <2.2e-16. **B.** Scatterplot showing mean DNA methylation values for genomic tiles in $Dnmt3a^{+/+}$ against $Dnmt3a^{+/D329A}$.

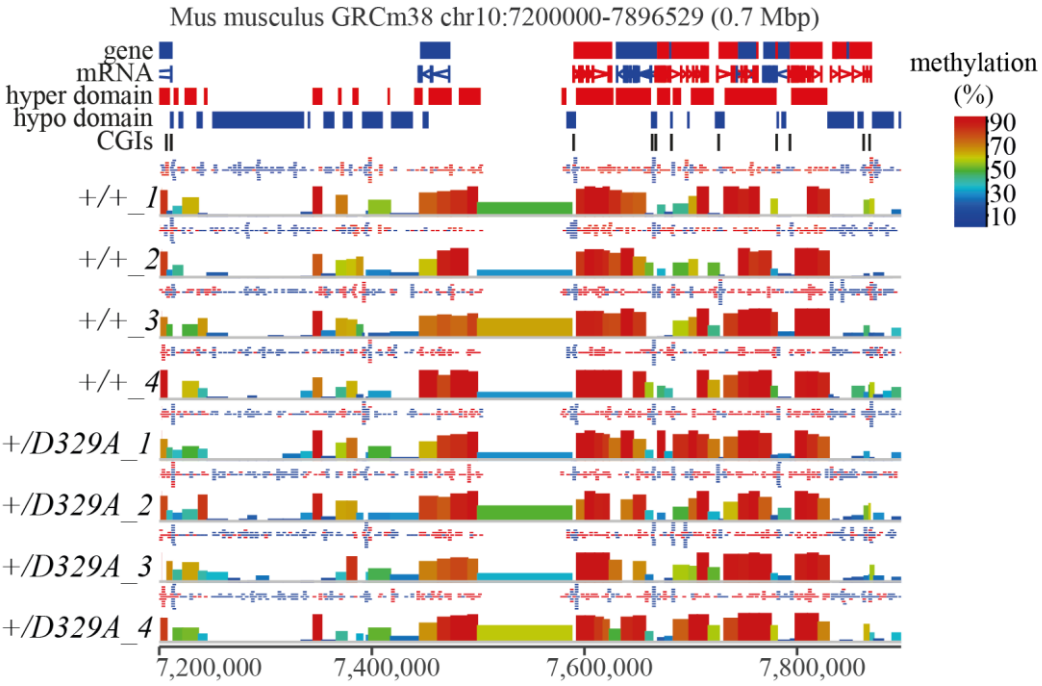


Figure 3.14. Genome browser view of oocyte pseudo-bulk samples. Alleles of *Dnmt3a* are provided as sample names. Each box represent 100CpG tile. Individual pseudo-bulk groupings are shown.

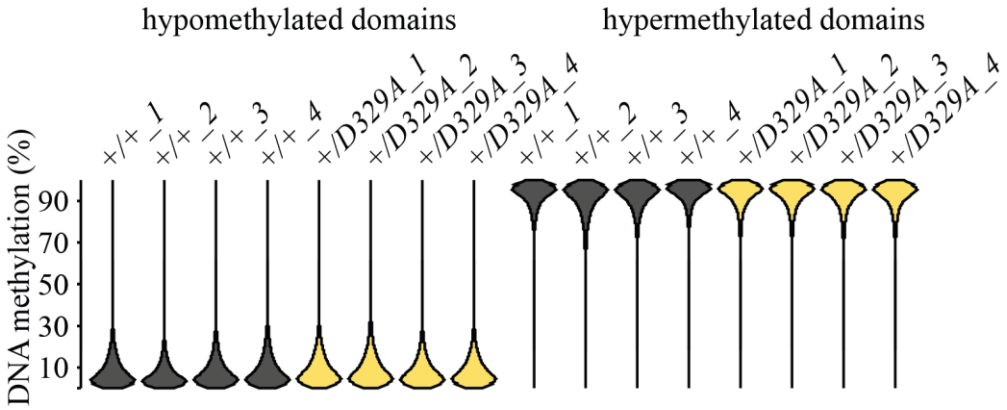


Figure 3.15. DNA methylation levels of hypo- and hyper- methylated domains in pseudo-bulk samples. 100CpG tiles were used. Wilcoxon rank sum test using mean tile values for each genotype results in p-value <2.2e-16 for hypomethylated and p-value = 0.0117 for hypermethylated domains.

DNA methylation in pseudobulk groups over regions of low enrichment (0-10% fraction) or high enrichment (90-100%) in histone marks H3K4me, H3K27me3 and H3K36me3 was assessed but did not show any apparent differences between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$ oocytes (Figure 3.16). Together these data agree with the findings of bulk oocyte analysis, and suggest that the D329A mutation does not cause functionally significant changes in genomic targeting of DNMT3A in mouse oocytes.

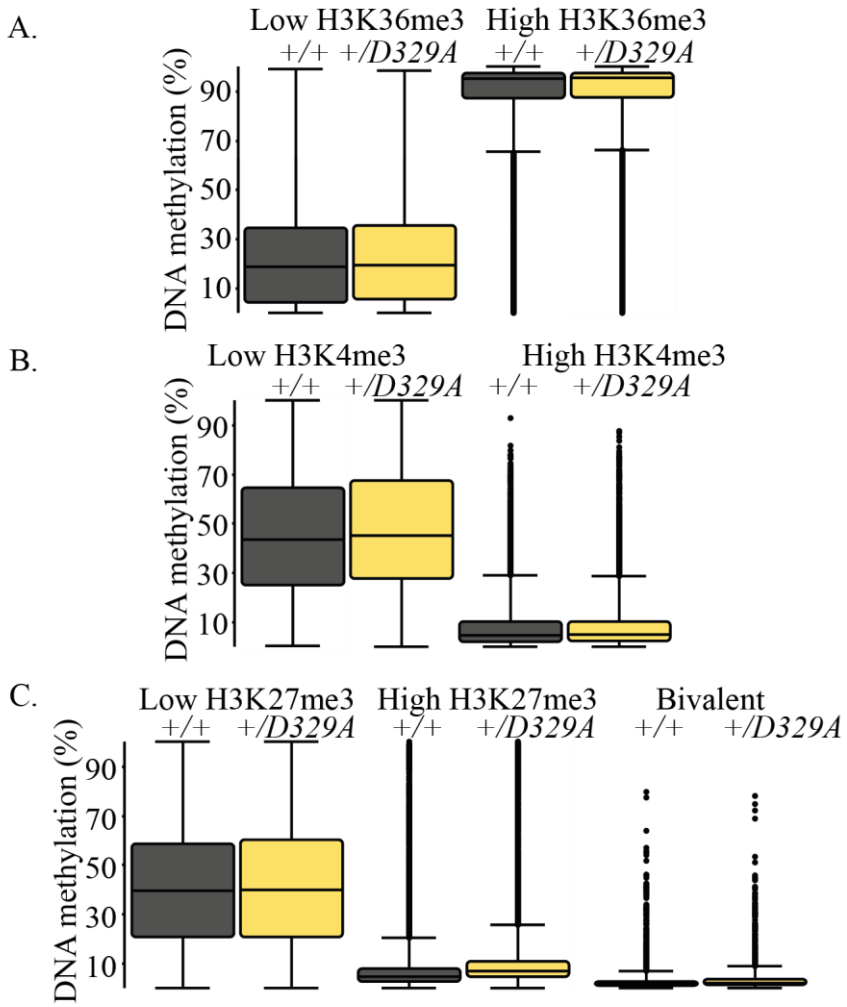


Figure 3.16. DNA methylation in pseudo-bulk oocytes samples over genomic domains linked to histone PTMs. High H3K4me3 and high H3K27me3 datasets exclude bivalent tiles. Low refers to bottom 10 % and high to top 10% of enrichment of genome-wide PTM distribution. Wilcoxon rank sum statistical test p-values listed in order of PTM domains, as they appear in the figure: **A.** p=0.054, p=0.0005, **B.** p=0.052, p=8.135e-15, **C.** p=0.754, p<2.2e-16, p<2.2e-16.

3.5 Summary and discussion

In this chapter, I assessed the impact of the presence of the *Dnmt3a*^{D329A} allele in mouse oocytes. I found that the D329A mutation does not dramatically affect the DNA methylation pattern in the oocyte. This is true over all genomic features and domains marked by varying levels of H3K4me3, H3K27me3, and/or H3K36me3. On the suspicion that there were varying levels of somatic cell and/or DNA contamination in bulk PBAT datasets, these findings were confirmed using a single-cell approach.

Redundancy between DNMT3A and DNMT3B makes it challenging to find tissues that would allow the dissection of the functionality of only one of these proteins. The mouse oocyte is one of those special cases where DNMT3A is known to be solely responsible for establishing DNA methylation and at no point during oocyte development is DNMT3B required (Kaneda et al., 2004, 2010). Given this, it was chosen as the first place for assessment of the impact of the D329A mutation. It came as a great surprise that no effect could be observed. There is a clear linear relationship between the key factors in *de novo* DNA methylation in the oocyte: Transcription, SETD2 catalysed H3K36me3, and DNMT3A established DNA methylation.

The first part of the hierarchy is transcription. Once a non-growing oocyte is induced to enter the growth phase, a definitive oocyte transcriptome is established. From around the mid-stage of oocyte growth, expressed genes start to acquire DNA methylation across their gene bodies (Gahurova et al., 2017). Methylation increases during the final stages of oocyte growth and is completed by the fully-grown GV stage. High transcriptional levels and DNA methylation show high correlation, and approximately 90% of methylome establishment has been attributed to transcription events (Gahurova et al., 2017; Kobayashi et al., 2012; Veselovska et al., 2015). While transcriptional activation and DNA methylation go hand in hand, it is unlikely that gene body DNA methylation is recruiting transcriptional units to the genome, and therefore it is unlikely that mutation in DNMT3A would have a direct effect on gene expression.

High transcriptional levels are not only linked to DNA methylation but also to H3K36me3 (Gahurova et al., 2017; Stewart et al., 2015; Veselovska et al., 2015; Xu et al., 2019). H3K36me3 is present in early oocyte growth stages and increases together with transcription and DNA methylation during oocyte maturation, persisting until the MII stage (Xu et al., 2019). Depletion of H3K36me3 in non-growing oocyte results in

dramatic loss and redistribution of DNA methylation and affects all maternal imprints in fully-grown oocytes. It also alters the transcriptome, leads to reorganisation of H3K4me3 and H3K27me3 mark landscape, and causes embryonic lethality (Xu et al., 2019). Bridging transcriptional activation and H3K36me3 is the enzyme SETD2 responsible for catalysing the H3K36me3 mark. SETD2 is bound to the RNA Polymerase II elongation subunit and is therefore able to set H3K36me3 over gene bodies (Yoh et al., 2008). Added together, findings from previous studies would strongly suggest that H3K36me3 is the key regulator of the oocyte epigenome, including DNA methylation. The only currently known way for DNMT3A to be recruited by H3K36me3 is via the PWWP domain (Dhayalan et al., 2010).

The isoform required for DNA methylation in the oocyte is the shorter DNMT3A2 isoform. It lacks the long intrinsically disordered part of the N-terminus, but retains both PWWP and ADD domains. The essential co-factor DNMT3L is also structurally similar to DNMT3A, but has the ADD domain only (Bourc'his et al., 2001). Given that the short isoform of DNMT3A is conferred by an alternative promoter, and many genes in oocytes have oocyte-specific promoters (Veselovska et al., 2015), we can speculate that the PWWP domain would be included in the shorter isoform transcript only if it conveyed a function in the oocyte.

If the mutation indeed disrupted the interaction between the DNMT3A-PWWP domain and H3K36me3, then we would expect to see a loss of DNA methylation in the oocyte domains marked by high levels of H3K36me3. There are three possible interpretations for why I do not detect any gross change: 1) the PWWP domain is not essential for recruitment of DNMT3A to the H3K36me3 marked genome; 2) the PWWP-D329A mutation does not disrupt the interaction *in vivo*; 3) DNMT3B is able to rescue the defect in the DNMT3A-PWWP domain, for example by heterodimerisation and targeting through the DNMT3B-PWWP domain.

The first possibility that the DNMT3A-PWWP domain is not required in oocytes would leave a question of why cellular energy and nucleic acid resources would be used to transcribe it in the oocyte isoform at all. The evidence for the role of H3K36me3 in DNA methylation recruitment is so compelling (Xu et al., 2019) that it would be difficult to imagine that another system, rather than a conserved PWWP domain, could be used. Could it be just because the PWWP domain is required in a maternal transcript in early

embryogenesis? Even so, this would not answer the question of how DNMT3A is recruited to genomic domains destined to be methylated.

A second possibility, that the PWWP-D329A mutation in DNMT3A does not have the same effect *in vivo* as *in vitro,* is an interesting speculation. One way to test it would be to delete the PWWP domain in the oocyte and look at DNA methylation levels. This experiment could provide an answer also to the question of whether the PWWP domain has a role in the oocyte at all.

Finally, a third possibility is of DNMT3B stepping in in the time of need. DNMT3A and DNMT3B are found in a functional complex in ESCs (Li et al., 2007a). *Dnmt3b* is expressed in the oocyte but deletion of it does not affect DNA methylation or the potential of the oocyte to be fertilised and form an embryo (Kaneda et al., 2010). While *Dnmt3b* is upregulated following *Dnmt3l* deletion (Lucifero et al., 2007), it does not appear to save the loss of DNA methylation in the *Dnmt3l*$^{-/-}$ mouse. Therefore, it is unlikely that a redundancy rescue mechanism was triggered by the D329A mutation in DNMT3A-PWWP domain in the oocyte.

In summary, this chapter shows that upon introduction of the D329A mutation in the DNMT3A PWWP domain, DNA methylation in the oocyte is not affected. Thus, the role of the PWWP domain in DNMT3A established *de novo* DNA methylation in the oocyte remains unclear. The negative result in particular highlights potential discrepancies of *in vitro* and *in vivo* studies, where the same mutation does not appear to cause the same outcome.

4 Mice carrying the *Dnmt3a*-D329A mutation exhibit a growth retardation phenotype

Declaration of author contributions

Mice weights were recorded by the Biological Support Unit staff at Babraham Institute. RNA-seq libraries were generated by Elena Ivanova.

An earlier version of this chapter has been published:

Sendžikaitė, G., Hanna, C.W., Stewart-Morgan, K.R., Ivanova, E., Kelsey, G., (2019). A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. Nature Communications, 10, Article 1884. https://doi.org/10.1038/s41467-019-09713-w

Mice carrying the D329A mutation in the DNMT3A-PWWP domain were initially generated to assess DNA methylation in the oocyte, where the only active *de novo* methyltransferase is DNMT3A, and the DNA-methylated genome is marked by H3K36me3. Although there were no changes found in oocytes of *Dnmt3a*^{D329A} mice, an observation was made that mice carrying the D329A mutation were much smaller than their littermates. This phenotype was unexpected, given the initial hypothesis that D329A mutation would have a loss-of-function outcome and because of the lack of phenotype in the oocyte. Intrigued by this observation, I decided to investigate the growth phenotype and underlying pathology.

4.1 Postnatal growth retardation phenotype

Initially, a serendipitous observation was made that littermate mice in the colony carrying the D329A mutation had varying sized bodies and heads (Figure 4.1). To further assess this phenotype, weights of mice from litters of four possible genotypes: *Dnmt3a*$^{+/+}$,

Dnmt3a$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$, *Dnmt3a*$^{\Delta/D329A}$ were recorded (Figure 1.6). Depending on availability, 4-19 pups per sex, per genotype were weighed at day 10, week 4, week 6, week 8, and week 12 to follow their growth patterns (Figure 4.2, Table 6). The weights were kindly recorded by the Biological Support Unit at Babraham Institute.

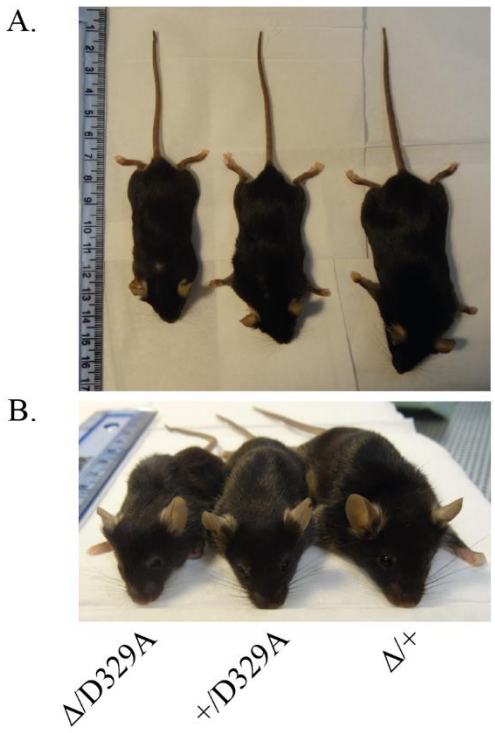


Figure 4.1. Photographs of littermate mice. A. Photograph showing whole body differences of littermate mice. **B.** Photograph showing head size differences of littermate mice. Alleles of *Dnmt3a* carried by mice are indicated. The ruler is in metric system, where one unit equals 1cm.

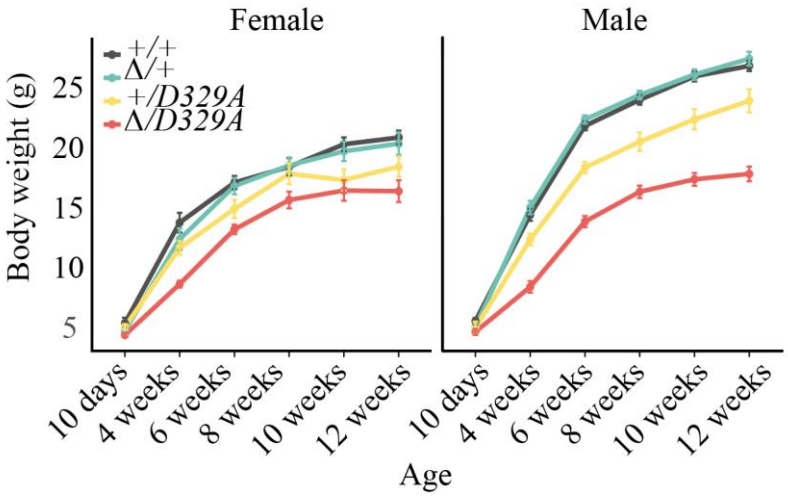


Figure 4.2. Body weights of mice recorded at different timepoints. Colours denote genotype, legend refers to *Dnmt3a* alleles. Replicate numbers are listed in Table 6. Error bars indicate error of the mean.

Table 6. Individual animals used (n value) in body weight curve generation at times indicated.

Male	**10 days**	**4 weeks**	**6 weeks**	**8 weeks**	**10 weeks**	**12 weeks**
+/+	19	19	18	19	19	18
Δ/+	12	12	12	12	12	12
+/D329A	7	7	8	8	7	7
Δ/D329A	12	12	12	13	13	11
Female	**10 days**	**4 weeks**	**6 weeks**	**8 weeks**	**10 weeks**	**12 weeks**
+/+	7	7	8	8	8	8
Δ/+	11	11	13	13	13	13
+/D329A	4	4	5	5	5	5
Δ/D329A	8	8	8	8	8	8

Dnmt3a$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$ mice displayed a postnatal growth retardation, which was more severe in males and in *Dnmt3a*$^{\Delta/D329A}$ animals. Mixed model ANOVA (KR-method) showed that the difference in weights was significantly dependent on genotype and time in both sexes (p<0.001, df(males) = 15,229.48, df(females) = 15,142.15), F=0). Subsequent pairwise Bonferroni-adjusted t-test comparisons showed that *Dnmt3a*$^{\Delta/D392A}$ females become significantly underweight starting from week 4 (p=0.002), with the exception of week 8 (p=0.095). *Dnmt3a*$^{+/D329A}$ females do not show statistically significant change. *Dnmt3a*$^{\Delta/D329A}$ males become significantly underweight from week 4 (p<0.001), and *Dnmt3a*$^{+/D329A}$ males from week 6 (p<0.001). *Dnmt3a*$^{\Delta/+}$ offspring were indistinguishable from *Dnmt3a*$^{+/+}$ in both sexes, indicating no haploinsufficiency of *Dnmt3a* at a phenotypic level. Notably, offspring carrying the *D329A* allele showed lower weights, even in the presence of a *wild-type* allele, suggesting D329A is a gain-of-function mutation.

In humans, dwarfism syndromes are sometimes caused by abnormal activation of the GH-IGF (growth hormone – insulin-like growth factor) axis (Ranke and Wit, 2018). GH-releasing hormone (GHRH) is secreted by the hypothalamus and activates anterior pituitary cells to release GH. GH directly promotes growth of bones and acts on the liver to activate IGF-I secretion. IGF-I enters the circulation and has a generic growth stimulation effect in most tissues, at the same time providing a negative feedback stimulus back to the pituitary gland.

I hypothesised that the growth restriction observed could have been caused by hormonal imbalance. To test this hypothesis, serum of 14-week old male mice was assessed using ELISA. At least four individual animals were used per genotype. However, the levels of GH and IGF-I were not significantly different across the four genotypes (p=0.9211 and p=0.1374, respectively) (Figure 4.3).

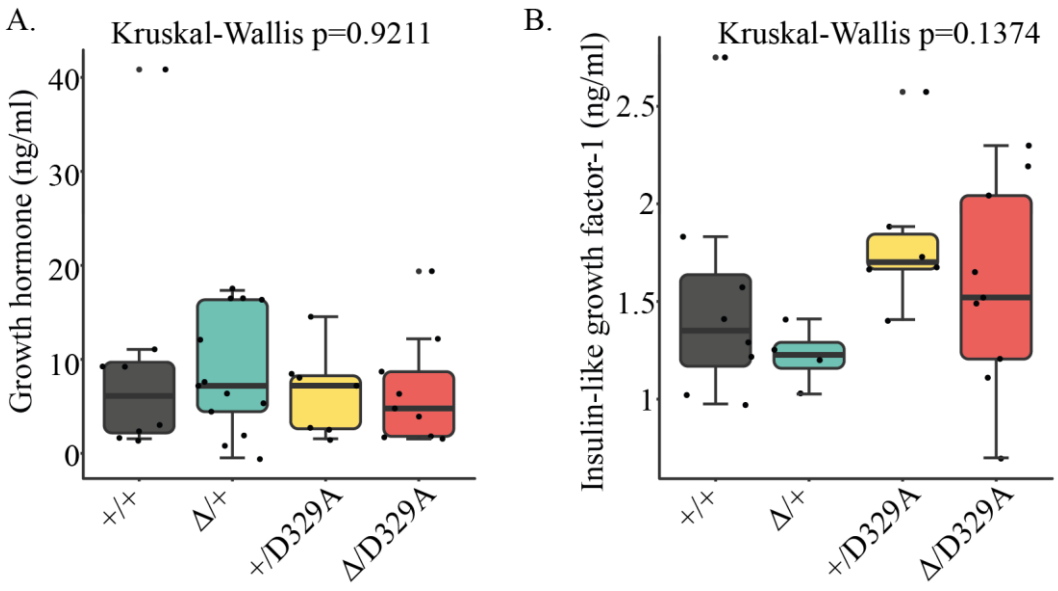


Figure 4.3. Levels of growth regulatory hormones in serum. Concentration of **A.** growth hormone and **B.** insulin-like growth factor-1 in serum of mice of indicated genotypes. Genotype refers to *Dnmt3a* alleles. Kruskal-Wallis is a non-parametric test used to assess statistical significance of the distribution. Each dot refers to one mouse.

4.2 Transcriptional changes in *Dnmt3a*^{D329A} mice

Dnmt3a is widely expressed in developing mouse brain starting from around embryonic day 11.5 (E11.5) to postnatal neurogenesis, and is the only *de novo* DNA methyltransferase expressed from E15.5 onward (Feng et al., 2005). To gain insights into causes of the abnormal growth RNA-seq was performed on adult (14-week) hypothalamus. The libraries were kindly generated by Elena Ivanova. Five replicates of *Dnmt3a*$^{+/+}$ and four replicates of *Dnmt3a*$^{\Delta/D329A}$ RNA-seq libraries were sequenced to the depth of 16.9-18.4 million reads and all passed FastQC checks. During analysis, only reads uniquely mapped to the genome were included. Quality control in the SeqMonk analysis tool indicated that most of the reads were overlapping gene annotations, and approx. 75% of all genes were covered (Figure 4.4). The libraries were generated using a directional opposite strand RNA-seq library kit and it was reassuring to observe almost all reads mapped on the anti-sense strand.

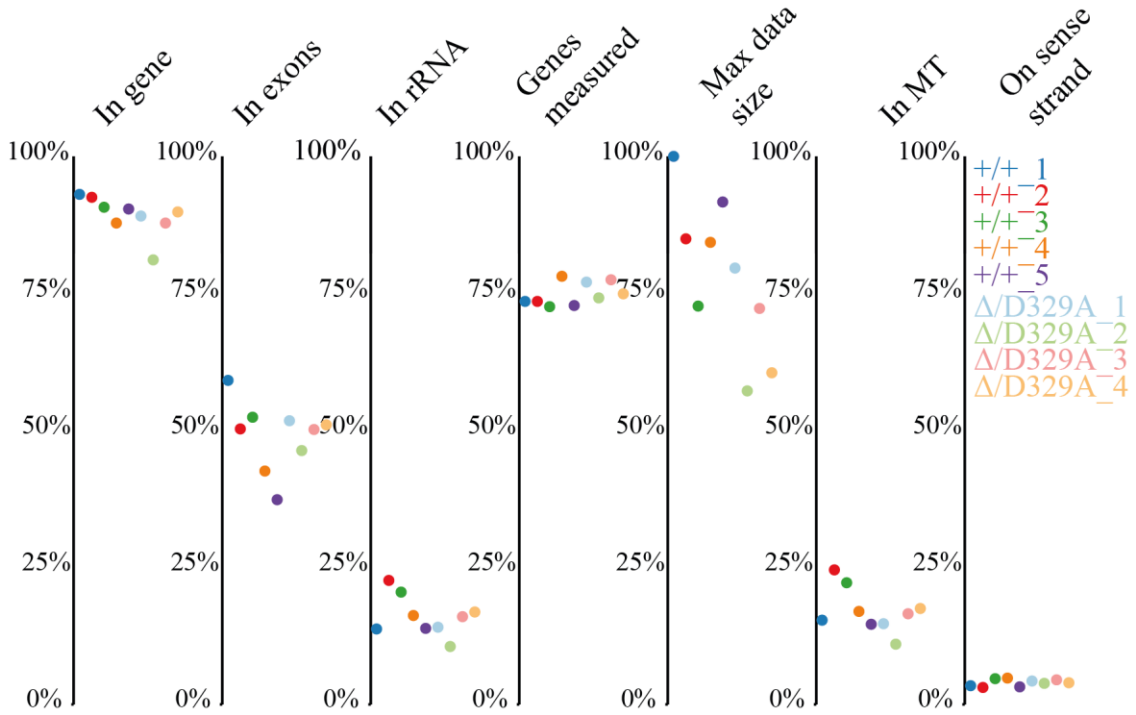


Figure 4.4. Quality control assessment of adult mice hypothalamus RNA-seq datasets. Graph shows a percentage of reads falling over genes, and exons in particular. Percentages of reads in rRNA or mitochondrial (MT) refer to levels of non-polyadenylated transcripts. Genes measured show how many genes would have a coverage of at least one read. Max data size show relative abundance of reads in all samples in reference to the largest one. On sense strand indicate how many transcripts fall onto sense strand.

First, expression levels of members of the DNA (de-)methylation machinery were assessed (*Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Dnmt3l*, *Tet1*, *Tet2*, and *Tet3*), but no statistically significant differences between $Dnmt3a^{+/+}$ and $Dnmt3a^{\Delta/D329A}$ animals were found (Figure 4.5). However, assessment of the whole transcriptome using DEseq (p<0.01, Benjamini-Hochberg multiple comparisons correction) identified 259 differentially expressed genes (DEGs) between the two genotypes (Figure 4.6). Upregulated genes fell mostly within the transcriptional regulator class, as defined by Gene ontology (GO) analysis (Figure 4.7). Interestingly, many normally strictly silenced developmental genes, such as *Tal1*, *Pou4f1*, *Wt1*, *Hoxa7*, and *Hoxd8*, exhibited upregulation in the $Dnmt3a^{\Delta/D329A}$ hypothalamus.

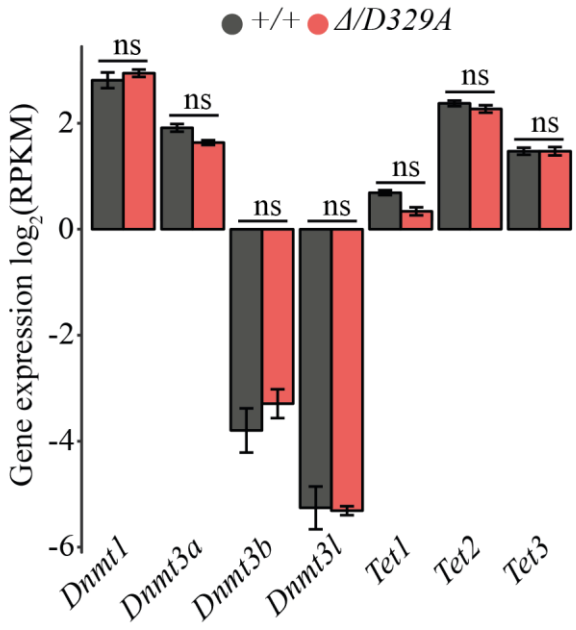


Figure 4.5. Expression of *Dnmt* and *Tet* family genes. Abundance of *Dnmt* and *Tet* gene transcripts in RNA-seq data from female adult hypothalamus. Two-tailed t-test was used with Bonferroni-corrected *p*-value<0.05 significance cut-off. ns.: not significant.

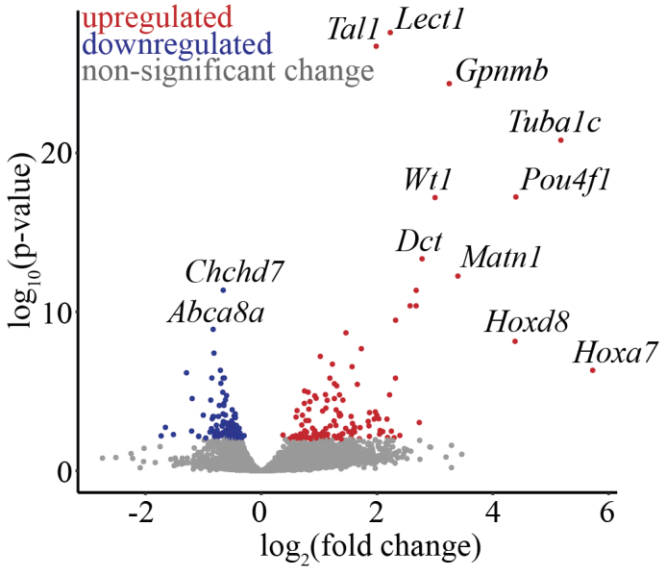


Figure 4.6. Differential gene expression. Volcano plot showing gene expression fold change and its significance between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ adult hypothalamus. Differentially expressed genes were determined using DEseq (p<0.01, Benjamini-Hochberg multiple comparisons correction). Selected genes showing the most significant and highest fold change are named. Colour denotes the status of a gene change in *Dnmt3a*$^{\Delta/D329A}$.

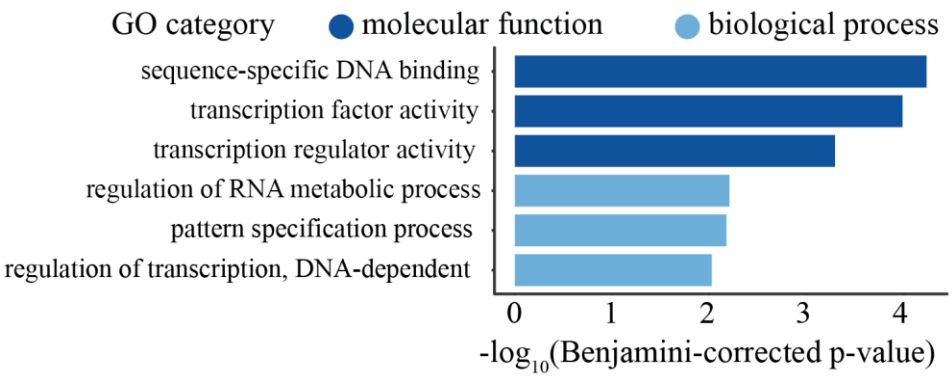


Figure 4.7. Gene ontology analysis of differentially expressed genes. Only significant terms are shown. Benjamini-Hochberg corrected *p*-value <0.01 was used as a threshold.

GO analysis did not find a significant enrichment of any functional subgroup within the downregulated genes. However, amongst those with decreased expression was the *Ghsr* gene, which encodes a growth hormone secretagogue receptor found in the hypothalamus and pituitary. It is activated by the circulating hormone ghrelin, which promotes a positive energy balance. Ghrelin stimulates food intake, increases adipogenesis and fatty acid storage, while decreasing energy expenditure (Mihalache et al., 2016). *Ghsr*$^{-/-}$ mouse show slightly but not significantly lower body weight than *wild-type* (Sun et al., 2008), suggesting decreased expression of this receptor could contribute to the growth phenotype, but is unlikely to be the key player.

4.3 Summary and discussion

Mice carrying a mutant *Dnmt3a*^{D329A} allele show a significant postnatal reduction in body weight. Growth restriction can also be observed in *Dnmt3a*$^{+/D329A}$ mice and presence of the *wild-type* allele does not rescue the phenotype, suggesting the D329A mutation has a dominant effect. The levels of GH and IGF-I were not significantly different between the genotypes; therefore, the cause of this defect is unlikely to be hormonal. Transcriptional analysis of adult hypothalamus unexpectedly showed transcriptional regulators as the key group with differential expression, namely, upregulation.

DNMT3A is known to be involved in regulation of growth, as *de novo* germline mutations in humans are found in Tatton-Brown-Rahman syndrome (TBRS, DNMT3A-overgrowth syndrome) (Tatton-Brown et al., 2014). Patients present with tall stature, macrocephaly and intellectual disability. Approximately 50 different mutations have been found in TBRS patients, varying from frameshift, missense, indels, stop-gain and whole gene deletions and similarly distributed across the whole protein: in intrinsically disordered N-term domain, PWWP domain, ADD domain and MTase domain (Tatton-Brown et al., 2018). *Dnmt3a*$^{+/-}$ heterozygous mice were originally described as being viable, fertile and having no immediately noticeable phenotype (Okano et al., 1999). A heterozygous *Dnmt3a* model has been recently revisited and a resemblance to TBRS was identified; with mice showing overgrowth tendencies and behavioural phenotypes (Christian et al., 2020). Nonetheless, these phenotypes had an adult-onset and the dosage effect appears to be stronger in humans.

Gain-of-function mutations in DNMT3A were discovered in three subjects with microcephalic dwarfism (Heyn et al., 2019). The mutations are located to the region orthologous to D329A in mouse and resulted in microcephalic dwarfism and growth restriction. The authors went on to generate a mouse model carrying the W326R mutation, which is only two amino acids away from the one described in this thesis. Similarly, both our and their mouse models recapitulate growth restriction observed in humans. Patients were also tested for GH and IGF-1 levels, which were not affected, in agreement with my findings in *Dnmt3a*^{D329A} mice. They also found that the brain weight was significantly lower in *Dnmt3a*$^{+/W326R}$ mice. This was not investigated in *Dnmt3a*^{D329A} mice.

One key difference between human *DNMT3A-PWWP* and *Dnmt3a*^{D329A} mouse pathologies was the onset of growth restriction; while in mice it manifested postnatally,

in patients with *DNMT3A-PWWP* mutations growth restriction was evident *in utero* (Heyn et al., 2019). Meanwhile, in 84% of TBRS patients, the overgrowth symptom is postnatal, similarly as observed in our *Dnmt3a*^{D329A} mouse model. This parallel indicates that DNMT3A might have a significant role in regulation of growth in mammals, but some differences in timing exist between mice and humans. Overall, findings in DNMT3A-associated pathology suggest that loss-of function mutations in DNMT3A lead to overgrowth, while gain-of-function mutations lead to growth restriction phenotypes, both in humans and mice, although the mechanism by which DNMT3A is involved in this regulation remains unclear.

Analysis of transcription in hypothalamus showed very few misregulated genes, given the extent of the phenotype. As a technical limitation, this could be attributed to a highly heterogeneous cell population in the hypothalamus. Based on transcriptomic patterns, adult mouse hypothalamus has 45 different cell type clusters, of which 34 are neuronal (Chen et al., 2017). Therefore, the effects on expression changes in a subset of them could easily be masked by bulk assessment. Nonetheless, it was surprising to observe that the targets of altered transcription are normally silenced developmental transcription factors. It is likely that the phenotype is caused by multi-genic action, which makes it rather difficult to unpick direct causes. Developmental genes are activated in a tightly controlled spatiotemporal manner during development; the consequences of expression in an adult tissue could be detrimental. On the other hand, expression of these genes could be an indirect effect of DNMT3A on DNA methylation. It is possible that changes, such as loss or gain of DNA methylation, would alter the local chromatin environment and expression changes would be a consequence. This is further investigated in the next chapters.

To sum up, here I show that presence of the *Dnmt3a*^{D329A} allele causes a dominant postnatal growth retardation in mice. Affected mice do not have detectably altered GH or IGF-1 levels, nor do they have altered levels of the DNA methylation or demethylation machinery. The cause of the growth restriction phenotype might be related to the de-repression of developmental regulator genes, which is observed in *Dnmt3a*^{D329A} adult hypothalamus.

5 The D329A mutation causes aberrant gain of methylation in adult tissues

Declaration of author contributions

Elena Ivanova kindly generated hypothalamus and pituitary PBAT libraries, Jessica Elmer and Wendy Dean collected E7.5 epiblasts.

An earlier version of this chapter has been published:

Sendžikaitė, G., Hanna, C.W., Stewart-Morgan, K.R., Ivanova, E., Kelsey, G., (2019). A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. Nature Communications, 10, Article 1884. https://doi.org/10.1038/s41467-019-09713-w

In Chapter 4, I described the postnatal growth restriction phenotype found in mice carrying the D329A mutation in the *Dnmt3a* gene. These mice exhibit de-repression of developmental transcription factors in the hypothalamus. Since DNMT3A is a DNA methyltransferase, I wanted to assess whether the methylome of mice carrying the D329A mutation is intact and whether changes in DNA methylation could be linked to altered gene expression.

5.1 DNA methylation in *Dnmt3a*^{D329A} mice

Genome-wide methylome maps were generated by PBAT sequencing in adult (14-week) hypothalamus from mice of all four genotypes to determine whether the D329A mutation had an impact on DNA methylation targeting by DNMT3A. Libraries were kindly generated by Elena Ivanova. Two to three replicates were generated per genotype. There were between 26.7 and 70.9 million reads per sample uniquely mapped and duplication rates varied between 3.6% and 6.4% (Table 7). All samples passed FastQC checks.

Table 7. Hypothalamus PBAT library details.

Sample ID	Duplication rate	Uniquely mapped and deduplicated alignments (million)
+/+_1	4.8%	41.9
+/+_2	5.1%	62.8
+/+_3	6.4%	46.8
Δ/+_1	3.6%	26.7
Δ/+_2	4.4%	34.3
Δ/+_3	4.8%	39.2
+/D329A_1	4.3%	58
+/D329A_2	5.9%	70.9
Δ/D329A_1	3.9%	39.5
Δ/D329A_2	4.4%	42.5
Δ/D329A_3	4.3%	42.5

The genome was binned into 100CpG tiles, and only those with at least 10 CpGs covered were used in the analysis. The global DNA methylation levels were similar across the four genotypes, at 77.2%, 75.1%, 77.2% and 75% in *Dnmt3a*$^{+/+}$, *Dnmt3a*$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329}$, respectively (Figure 5.1). The 2.1% lower global DNA methylation level in *Dnmt3a*$^{\Delta/+}$ compared to *Dnmt3a*$^{+/+}$ suggested a degree of haploinsufficiency. Similarly, *Dnmt3a*$^{\Delta/D329}$ showed a 2.2% lower global methylation compared to *Dnmt3a*$^{+/D329A}$, suggesting ~2% of the genome-wide DNA methylation in the hypothalamus is sensitive to *Dnmt3a* dosage. Although global DNA methylation levels were so similar, PCA analysis showed spatial separation of samples (Figure 5.2). Replicates coming from the same genotype clustered together, but there was a distinct

separation of different genotypes. PC1 separation appeared to be driven by the presence of a mutant allele, whilst PC2 separation was likely driven by the number of alleles expressed. This effect was highly reproducible between animals.

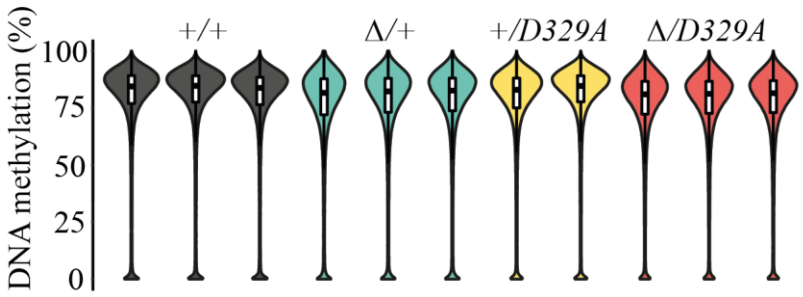


Figure 5.1. Global DNA methylation in adult mice hypothalamus. Colour denotes genotype, alleles indicate the status of *Dnmt3a*. Individual samples represent one mouse. 100 CpG tiles were used. Kruskall-Wallis rank sum test using mean tile values for each genotype results in p-value <2.2e-16 and is significant for all post-hoc pairwise Wilcoxon rank sum tests.

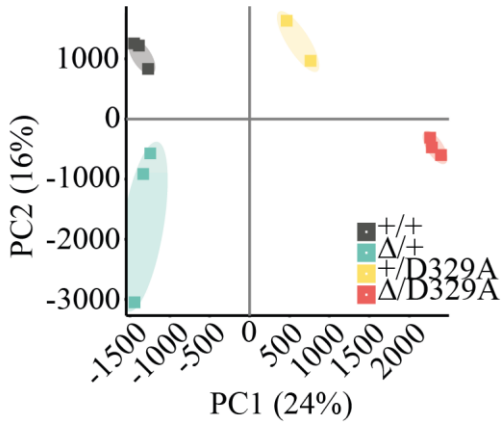


Figure 5.2. PCA analysis of hypothalamus samples. Colour denotes genotype. Alleles refer to *Dnmt3a* gene. 100 CpG tiles were used.

The PWWP domain of DNMT3A was predicted to target regions of H3K36me3 over the genome; the D329A mutation ablated this interaction *in vitro* (Dhayalan et al., 2010). Since the PCA plot indicated that samples with the D329A mutation had a distinct

methylome, I hypothesised that the D329A mutation would result in loss of DNA methylation over H3K36me3 domains. Since actively transcribed genes bodies are marked by H3K36me3, and hypothalamus RNA-seq datasets were available (Chapter 5), I started this investigation by looking at gene bodies. First, I assigned genes to four categories based on gene expression levels: non-expressed ($\log_2$(RPKM) < 0), low-expressed (0 < $\log_2$(RPKM) <1), mid-expressed (1 < $\log_2$(RPKM) <3), high expressed ($\log_2$(RPKM)>3); and then examined levels of DNA methylation over these domains. The levels of DNA methylation could be attributed to the global methylation differences observed across genotypes, which were not associated with gene expression pattern. Independently of their expression status or mouse genotype, gene bodies had relatively high DNA methylation of more than 68% (Figure 5.3). Within expression classes, the variation in mean methylation between genotypes was never more than 3.2%. Similar to the observed effect at the level of global DNA methylation, *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$ samples had approx. 2% lower DNA methylation than *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$, respectively.

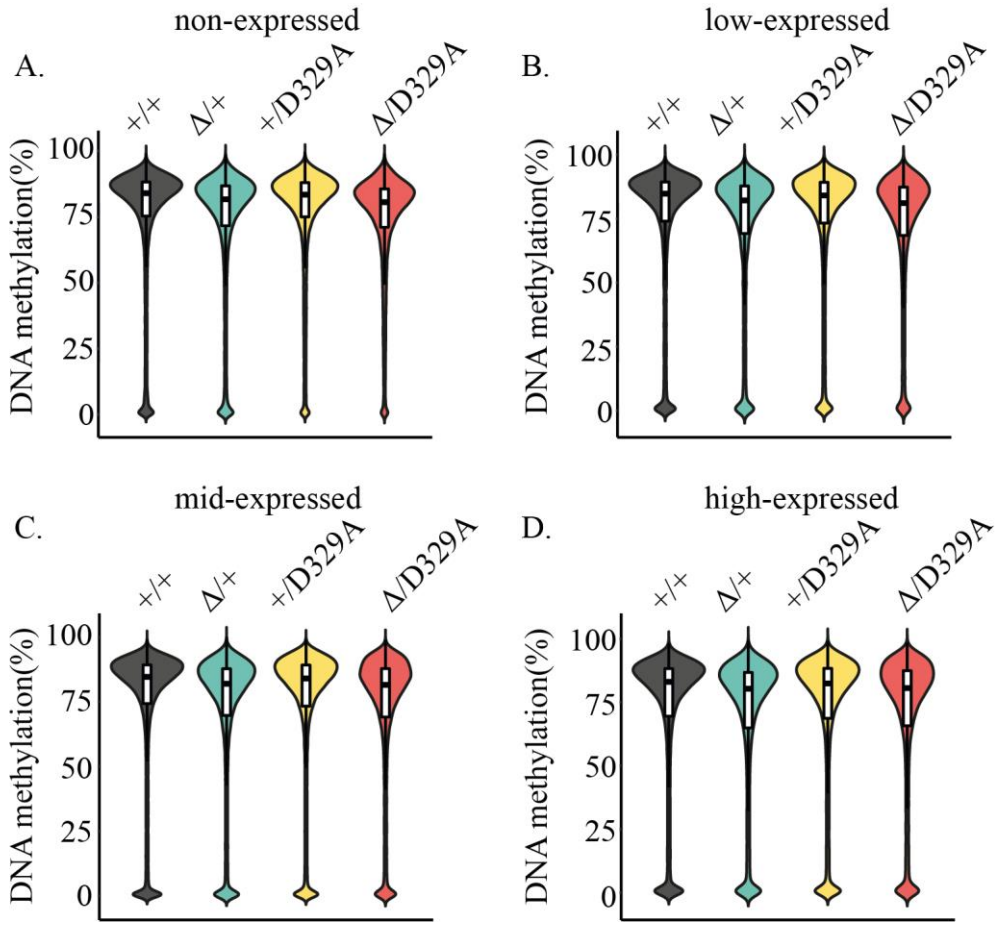


Figure 5.3. Gene body DNA methylation levels. DNA methylation levels over **A.** non-expressed, **B.** low-expressed, **C.** mid-expressed, **D.** high-expressed gene bodies. Definition of expression level is as follows: non-expressed: $\log_2(RPKM) < 0$; low-expressed: $0 < \log_2(RPKM) < 1$; mid-expressed: $1 < \log_2(RPKM) < 3$; high-expressed: $\log_2(RPKM) > 3$. Alleles indicated refer to *Dnmt3a* gene. Kruskall-Wallis rank sum test results in p-value <2.2e-16 and is significant for all post-hoc pairwise Wilcoxon rank sum tests. The only exceptions are in pairs of **C.** $Dnmt3a^{\Delta/+}$ and $Dnmt3a^{\Delta/D329A}$ Wilcoxon rank sum test p-value=0.36 and **D.** $Dnmt3a^{+/+}$ and $Dnmt3a^{+/D329A}$ Wilcoxon rank sum test p-value=0.9.

Next, I investigated DNA methylation over promoter regions, dividing them into those overlapping CGIs or not. Non-CGI promoters showed a pattern resembling global methylation trends, where $Dnmt3a^{+/+}$ and $Dnmt3a^{+/D329A}$ had approximately 81.8% DNA methylation, and $Dnmt3a^{\Delta/+}$ and $Dnmt3a^{\Delta/D329A}$ promoters had 2.2% lower DNA methylation (Figure 5.4a). However, at CGI promoters the pattern differed. Methylation of CGI promoters was bimodally distributed, with mean methylation of 38.7%, 36.9%,

40.4% and 39.9% in *Dnmt3a*$^{+/+}$, *Dnmt3a*$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329}$, respectively (Figure 5.4b).

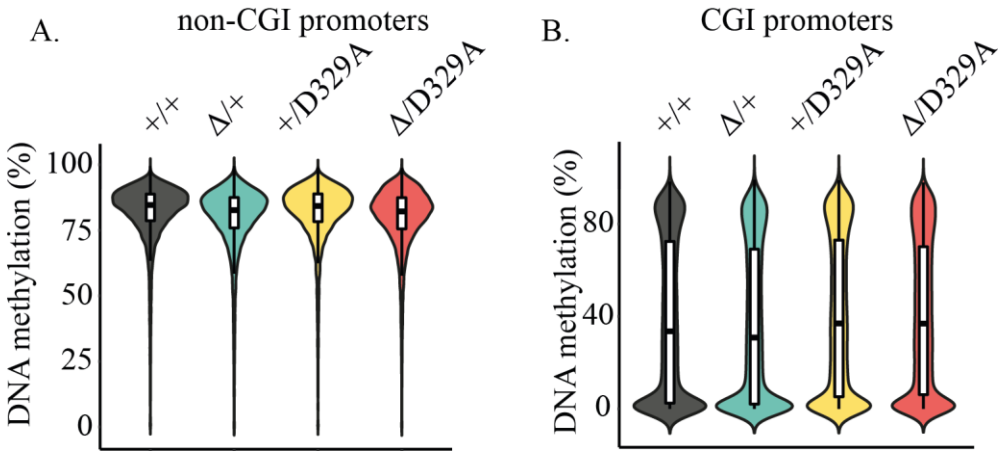


Figure 5.4. DNA methylation over promoter regions. Levels of DNA methylation over tiles overlapping **A.** non-CGI, and **B.** CGI promoters. Alleles refer to *Dnmt3a* gene, 100CpG tiles were used. Kruskall-Wallis rank sum test results in p-value <2.2e-16 and is significant for all post-hoc pairwise Wilcoxon rank sum tests. The only exceptions are pairs of **A.** *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$, and **B.** *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$, where Wilcoxon rank sum test p-value=0.19 and p-value=0.12, respectively.

Higher mean DNA methylation levels at CGI promoter regions in mice carrying the D329A mutation was a rather unexpected finding. To further assess this, I used the edgeR differential methylation tool to identify genomic tiles with significantly different DNA methylation levels between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ samples. It identified 4,184 differentially methylated genomic regions (DMRs) (p(adj)<0.01), which represent approx. 1% of the genome. Surprisingly, 87.5% of the DMRs were hypermethylated in *Dnmt3a*$^{\Delta/D329A}$ hypothalamus in comparison to *Dnmt3a*$^{+/+}$ (Figure 5.5). The level of DNA methylation over DMRs was consistent in all replicates, despite them originating from different mice (Figure 5.6). Identified hypermethylated DMRs were normally unmethylated, showing approx. 12% and 9.2% DNA methylation in *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/+}$ mice, respectively (Figure 5.6a), and methylated at an average level of 37.7% and 47.5% in *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$, respectively. Hypomethylated DMRs, on the other hand, were relatively highly methylated at approx. 73.1% in *Dnmt3a*$^{+/+}$ and 63%

in *Dnmt3a*$^{\Delta/+}$ hypothalamus, and showed a lower DNA methylation of 68.1% in *Dnmt3a*$^{+/D329A}$ and 52.4% in *Dnmt3a*$^{\Delta/D329A}$ hypothalamus (Figure 5.6b).

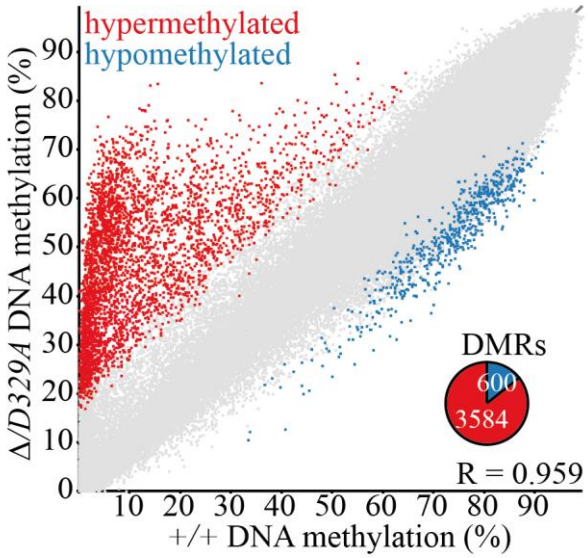


Figure 5.5. DNA methylation differences in *Dnmt3a*$^{+/+}$ compared with *Dnmt3a*$^{\Delta/D329A}$ hypothalamus. Scatterplot showing correlation of methylation levels of individual 100-CpG tiles between and *Dnmt3a*$^{\Delta/D329A}$ hypothalamus. Differentially methylated probes were determined using the edgeR proportion statistic in SeqMonk (p<0.01 corrected for multiple comparisons using Benjamini-Hochberg, methylation difference ≥20%). Pie-chart indicates how many of DMRs are hyper- or hypo- methylated in *Dnmt3a*$^{\Delta/D329A}$.

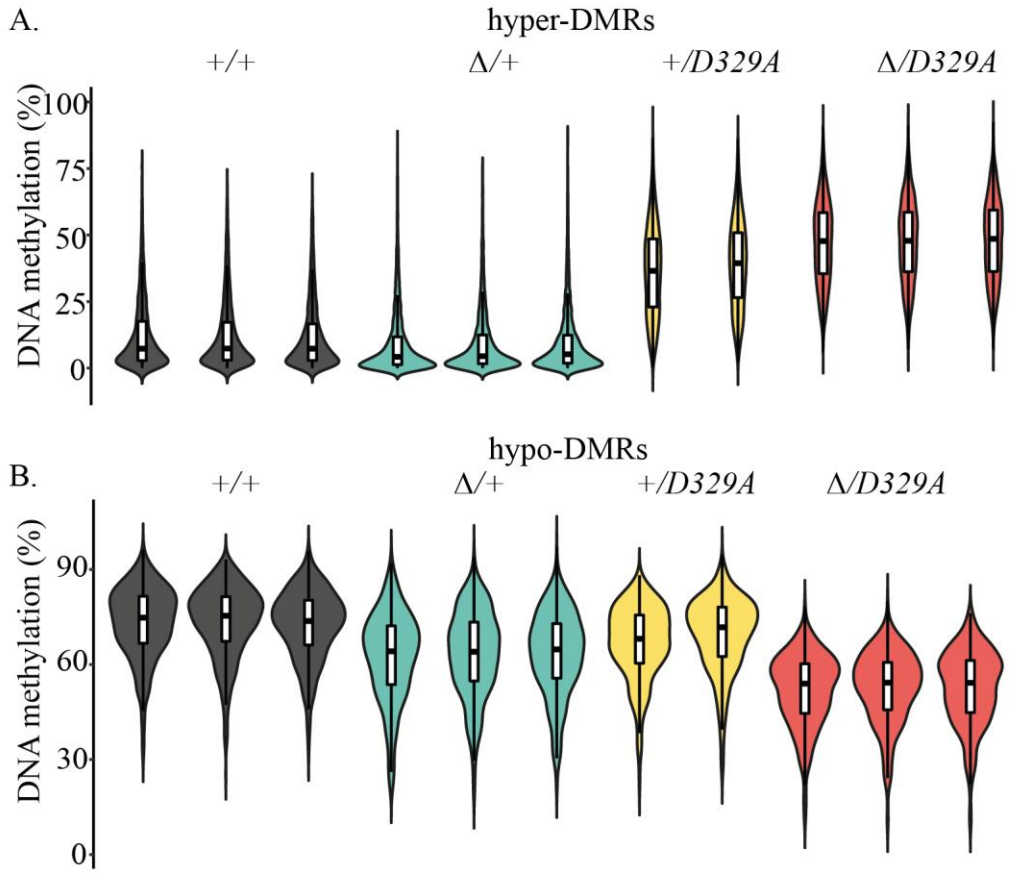


Figure 5.6. DNA methylation levels at DMRs in individual hypothalamus samples. Beanplots showing reproducible DNA methylation changes over (**a**) hyper- and (**b**) hypo-DMRs in 14-week adult male hypothalamus across four different genotypes. Each plot represents an individual mouse. Hypo: hypomethylated, hyper: hypermethylated, DMR: differentially methylated region. 100CpG tiles were used. Kruskall-Wallis rank sum test using mean tile values for each genotype results in p-value <2.2e-16 and is significant for all post-hoc pairwise Wilcoxon rank sum tests.

In order to dissect whether differential methylation was caused by the different number of *Dnmt3a* alleles or the D329A mutation, the edgeR tool was used to compare *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/+}$, and *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$ samples. As predicted, comparison of *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/+}$ yielded 351 DMRs, of which 328 were hypomethylated in *Dnmt3a*$^{\Delta/+}$ hypothalamus (Figure 5.7a). In contrast, comparison of *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$ samples, yielded 4603 DMRs, of which 4515 were hypermethylated (Figure 5.7b). These findings suggest that the hypermethylated DMRs are associated with the mutant *Dnmt3a*^{D329A} allele, while hypomethylated DMRs are linked to the number of *Dnmt3a* alleles and are a result of haploinsufficiency.

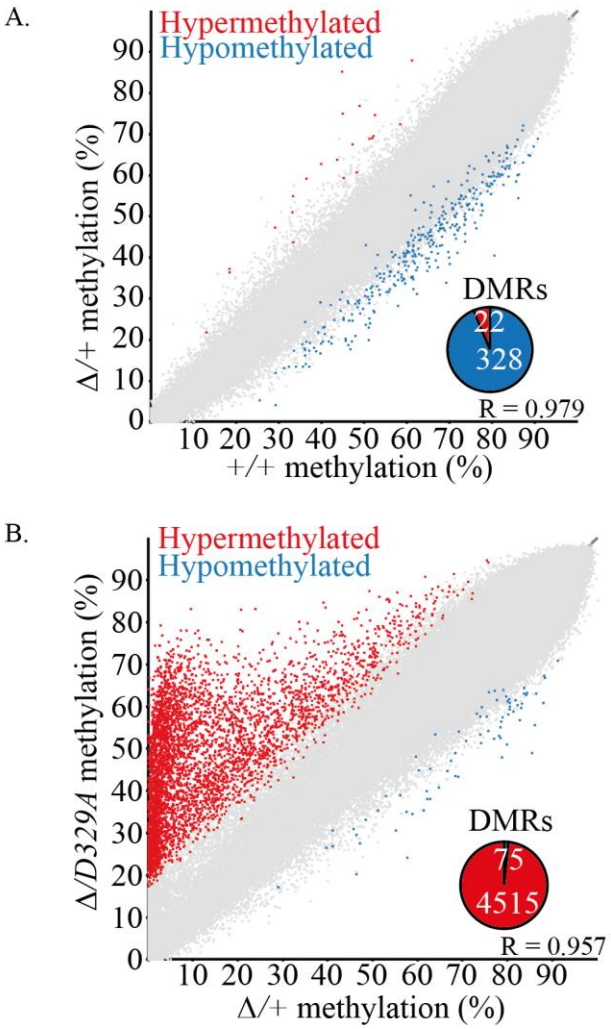


Figure 5.7. DNA methylation changes in adult hypothalamus between different genotypes. Scatterplots showing correlation between methylation levels of 100-CpG tiles in hypothalamus of **A.** *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/+}$ and **B.** *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$. Differentially methylated tiles were determined using EdgeR (p<0.01 corrected for multiple comparisons using Benjamini-Hochberg, methylation difference ≥20%). The pie-chart indicates the number of DMRs identified.

5.2 Link between DNA methylation and gene expression changes

Many genomic DMR tiles were found to be adjacent to each other and some covered areas spanning promoters or whole genes (Figure 5.8). Therefore, I set out to understand whether the changes in DNA methylation caused changes in gene expression found in Chapter 3. DEG regions were separated based on gene expression change, promoter and gene body DNA methylation and clustered to identify any patterns. Only approx. 20%

misregulated genes were linked to changes in DNA methylation, independently of whether they were up- or down-regulated in mutant mice, suggesting that a substantial proportion of DEGs may be attributable to indirect, downstream effects of direct targets.

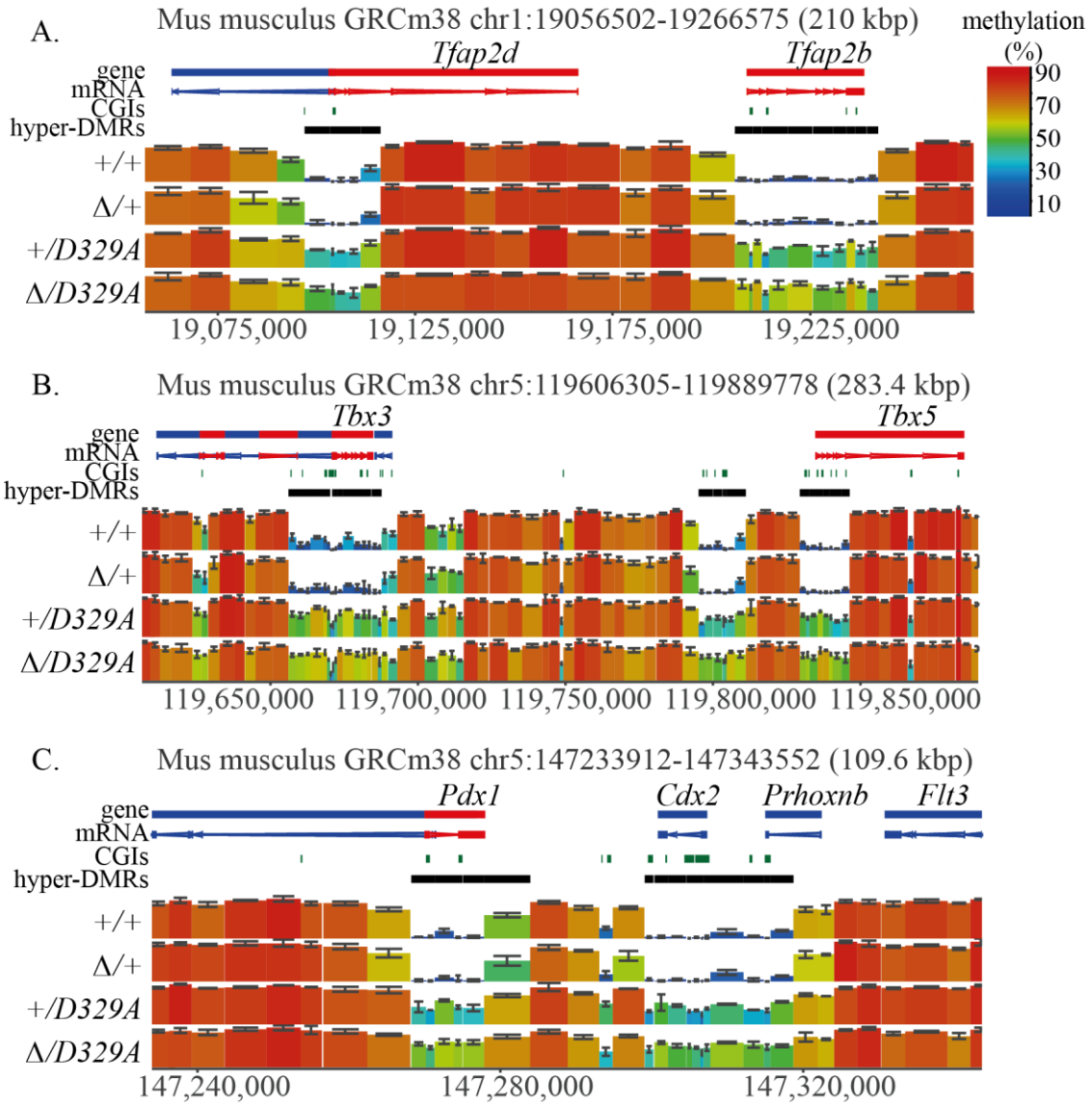


Figure 5.8. Genome browser views of representative genes that gain methylation. A., **B.**, and **C.**, show three representative affected loci. Methylation profiles for hypothalamus for each *Dnmt3a* genotype are shown. Each block, colour-coded for methylation value, represents a 100-CpG tile. Error bars indicate standard deviation. For the mRNA tracks, red indicates genes transcribed from left to right and blue indicates genes transcribed from right to left.

Amongst those genes with altered DNA methylation, two main groups could be identified using Euclidean clustering analysis: hypomethylated and hypermethylated promoter clusters (Figure 5.9 and Figure 5.10). These were evident in both upregulated and downregulated gene sets. Interestingly, almost all genes with hypermethylated promoters showed gain of DNA methylation over the gene body too (Figure 5.9, Figure 5.10, Figure 5.11 and Figure 5.12). This cluster contained the same developmental transcription factor genes that are normally strongly repressed in adult hypothalamus, but which are paradoxically derepressed in *Dnmt3a*$^{\Delta/D329A}$ (e.g., *Hoxa7, Hoxd8, Pou4f1, Tal1* and *Wt1*). This effect was weaker in the downregulated gene set. Genes with hypomethylated promoters did not show much change in gene body methylation.

Overall, a broad gain of DNA methylation over the gene could result from increased expression if the promoter remained relatively unmethylated, as was the case for *Pou4f1* (Figure 5.13a), or when the promoter was methylated as well, as for *Lxh9* (Figure 5.13b). Conversely, the *Ghsr* gene, which had a methylated gene body in *Dnmt3a*$^{+/+}$, gained DNA methylation over the promoter region and showed decreased expression (Figure 5.13c). Taken together, these findings suggest that the *Dnmt3a*^{D329A} allele is a driver of aberrant gain of DNA methylation and derepression of a subset of associated genes.

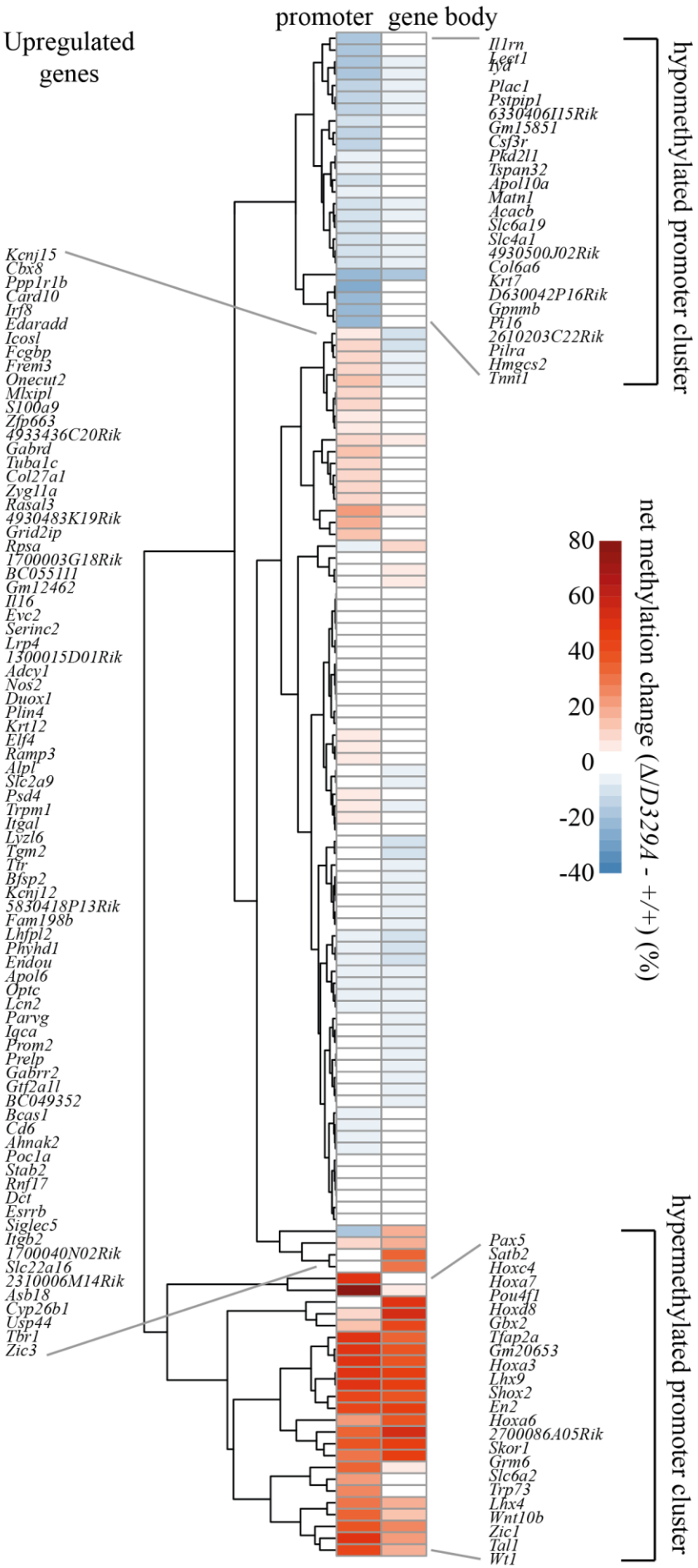

promoter
gene body
Upregulated genes
Kcnj15
Cbx8
Ppp1r1b
Card10
Irf8
Edaradd
Icosl
Fcgbp
Frem3
Onecut2
Mlxipl
S100a9
Zfp663
4933436C20Rik
Gabrd
Tuba1c
Col27a1
Zyg11a
Rasal3
4930483K19Rik
Grid2ip
Rpsa
1700003G18Rik
BC055111
Gm12462
Il16
Evc2
Serinc2
Lrp4
1300015D01Rik
Adcy1
Nos2
Duox1
Plin4
Krt12
Elf4
Ramp3
Alpl
Slc2a9
Psd4
Trpm1
Itgal
Lyzl6
Tgm2
Tfr
Bfsp2
Kcnj12
5830418P13Rik
Fam198b
Lhfpl2
Phyhd1
Endou
Apol6
Optc
Lcn2
Parvg
Igca
Prom2
Prelp
Gabrr2
Gtf2a1l
BC049352
Bcas1
Cd6
Ahnak2
Poc1a
Stab2
Rnf17
Dct
Esrrb
Siglec5
Itgb2
1700040N02Rik
Slc22a16
2310006M14Rik
Asb18
Cyp26b1
Usp44
Tbr1
Zic3
Il1rn
Lect1
Ly6
Plac1
Pstpip1
6330406I15Rik
Gm15851
Csf3r
Pkd2l1
Tspan32
Apol10a
Matn1
Acacb
Slc6a19
Slc4a1
4930500J02Rik
Col6a6
Krt7
D630042P16Rik
Gpnmb
Pi16
2610203C22Rik
Pilra
Hmgcs2
Tnni1
Pax5
Satb2
Hoxc4
Hoxa7
Pou4f1
Hoxd8
Gbx2
Tfap2a
Gm20653
Hoxa3
Lhx9
Shox2
En2
Hoxa6
2700086A05Rik
Skor1
Grm6
Slc6a2
Trp73
Lhx4
Wnt10b
Zic1
Tal1
Wt1
hypomethylated promoter cluster
hypermethylated promoter cluster
80
60
40
20
0
-20
-40
net methylation change (Δ/D329A - +/+) (%)

Figure 5.9. Methylation changes over promoters and gene bodies of upregulated DEGs. Net methylation differences between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D392A}$ were quantified over promoters and gene bodies. Promoter probes within 1kb distance were merged. Clustering was performed using the Euclidean complete method.

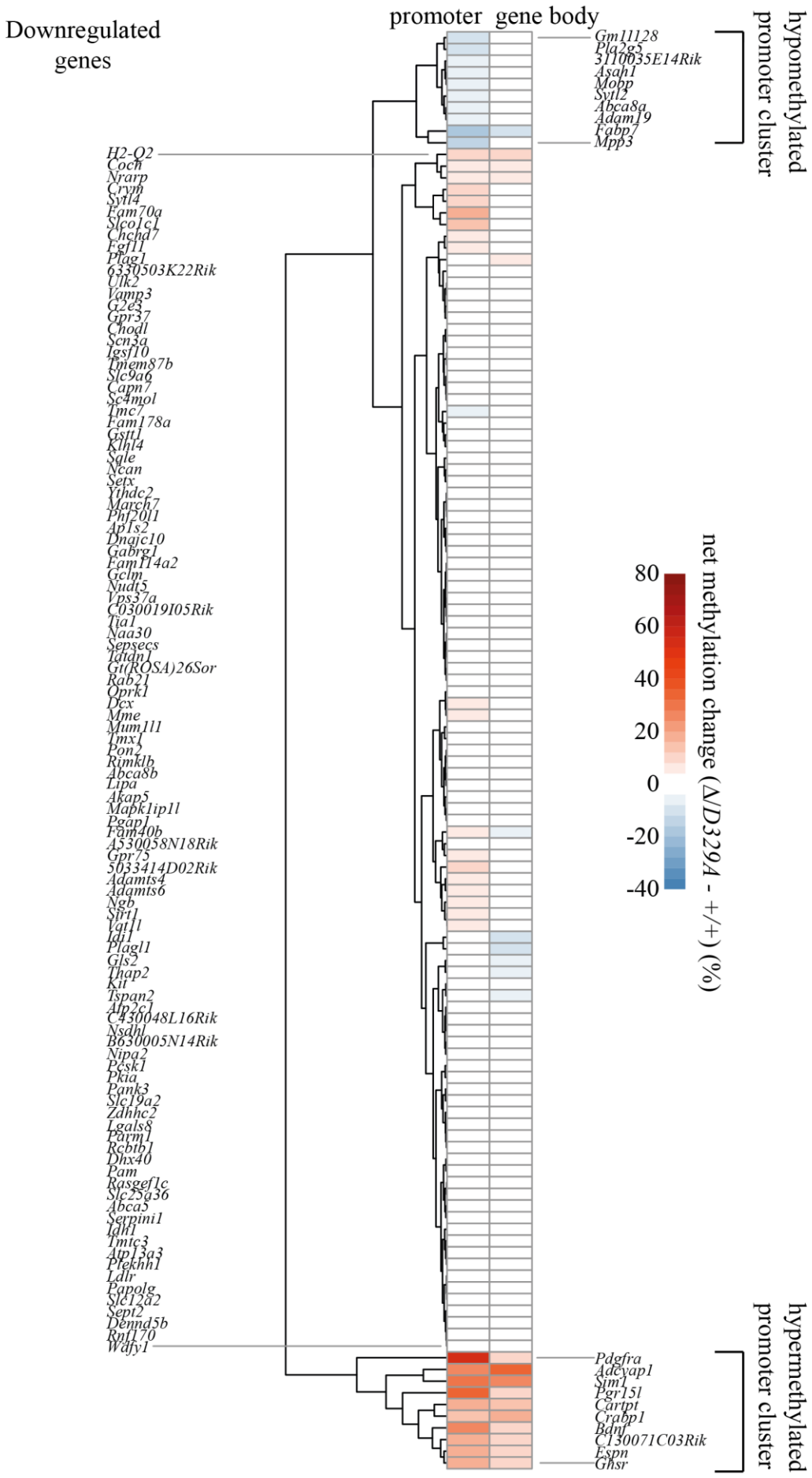

Downregulated genes
promoter
gene body
hypomethylated promoter cluster
hypermethylated promoter cluster
net methylation change (Δ*D329A* - +/+) (%)
80
60
40
20
0
-20
-40
Gm11128
Pla2g5
3110035E14Rik
Asah1
Mobp
Sytl2
Abca8a
Adam19
Fabp7
Mpp3
H2-Q2
Coch
Nrarp
Crym
Sytl4
Fam70a
Slco1c1
Chchd7
Fgf11
Ptag1
6330503K22Rik
Ulk2
Vamp3
G2e3
Gpr37
Chodl
Scn3a
Igsf10
Tmem87b
Slc9a6
Capn7
Sc4mol
Tmc7
Fam178a
Gstt1
Klhl4
Sqle
Ncan
Setx
Ythdc2
March7
Phf20l1
Ap1s2
Dnajc10
Gabrg1
Fam114a2
Gclm
Nudt5
Vps37a
C030019I05Rik
Tia1
Naa30
Sepsecs
Tdtdn1
Gt(ROSA)26Sor
Rab21
Oprk1
Dcx
Mme
Mum1l1
Tmx1
Pon2
Rimklb
Abca8b
Lipa
Akap5
Mapk1ip1l
Pgap1
Fam40b
A530058N18Rik
Gpr75
5033414D02Rik
Adamts4
Adamts6
Ngb
Slit1
Vat1l
Idi1
Plagl1
Gls2
Thap2
Kit
Tspan2
Atp2c1
C430048L16Rik
Nsdhl
B630005N14Rik
Nipa2
Pcsk1
Pkia
Pank3
Slc19a2
Zdhhc2
Lgals8
Parm1
Rcbtb1
Dhx40
Pam
Rasgef1c
Slc25a36
Abca5
Serpini1
Idh1
Tmtc3
Atp13a3
Plekhh1
Ldlr
Papolg
Slc12a2
Sept2
Dennd5b
Rnf170
Wdfy1
Pdgfra
Adcyap1
Sim1
Pgr15l
Cartpt
Crabp1
Bdnf
C130071C03Rik
Espn
Ghsr

Figure 5.10. Methylation changes over promoters and gene bodies of downregulated DEGs. Net methylation differences between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D392A}$ were quantified over promoters and gene bodies. Promoter probes within 1kb distance were merged. Clustering was performed using the Euclidean complete method.

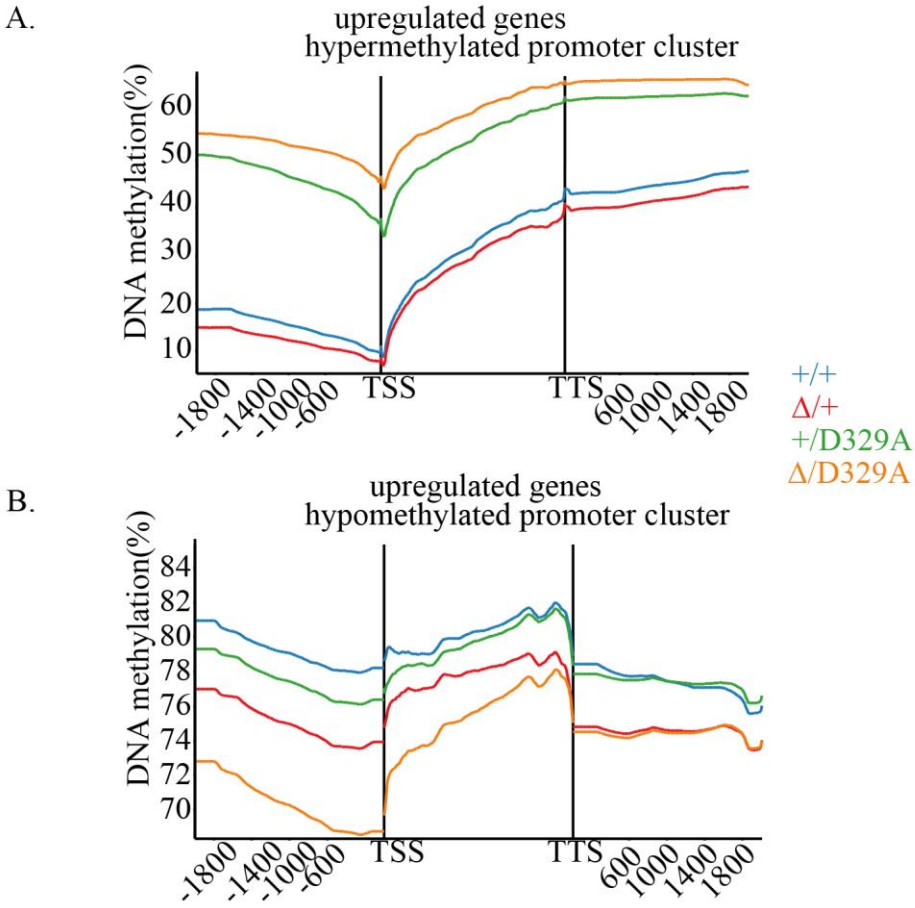


Figure 5.11. DNA methylation trendplot over upregulated genes. Aggregated methylation levels over upregulated gene clusters showing hyper- and hypo- methylated promoters, marked in Figure 5.9. TSS: transcription start site, TTS: transcription termination site. Distance over gene is relative.

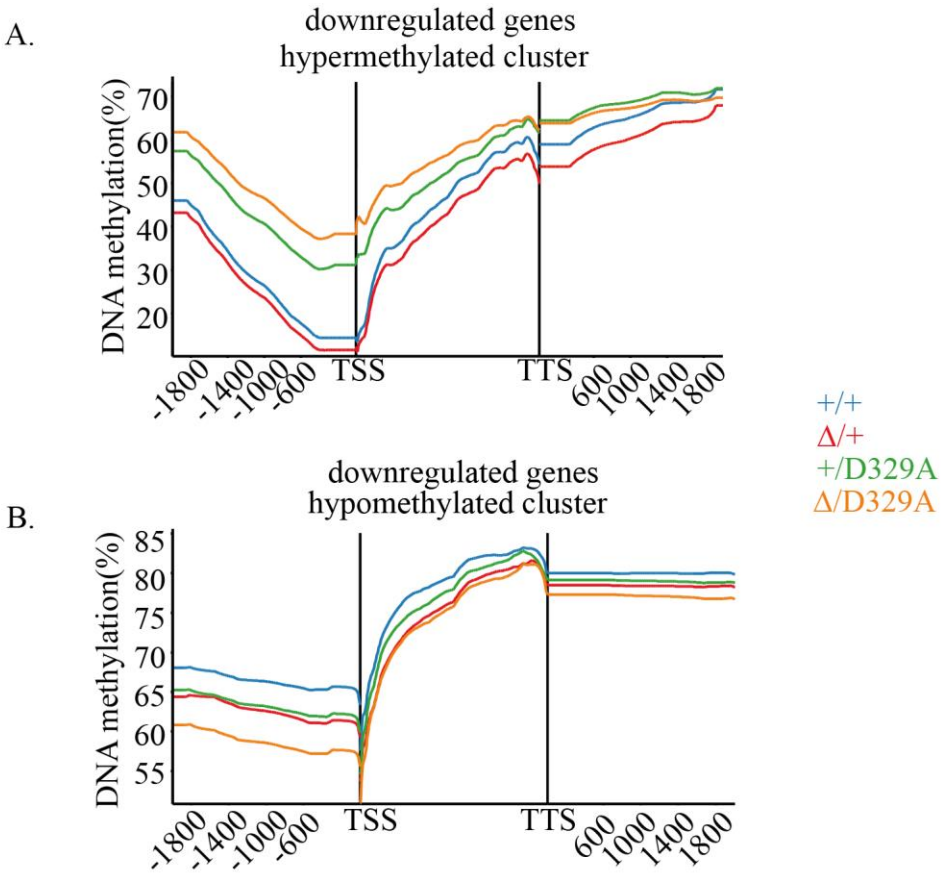


Figure 5.12. DNA methylation trend over downregulated genes. Aggregated methylation levels over upregulated gene clusters showing hyper- and hypo- methylated promoters, marked in Figure 5.10. TSS: transcription start site, TTS: transcription termination site. Distance over gene is relative.

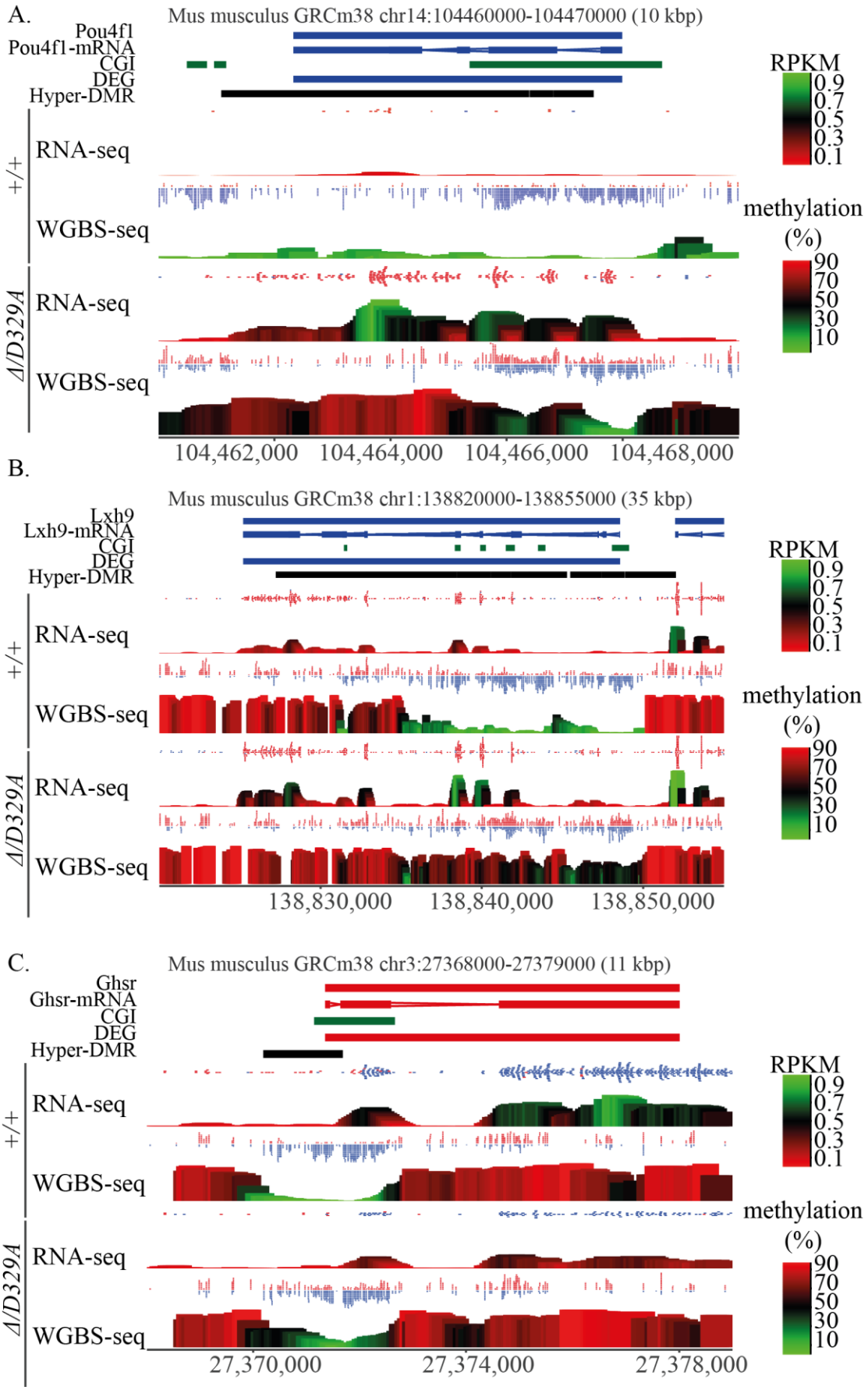

A.
Mus musculus GRCm38 chr14:104460000-104470000 (10 kbp)
Pou4f1
Pou4f1-mRNA
CGI
DEG
Hyper-DMR
+/+
RNA-seq
WGBS-seq
Δ/D329A
RNA-seq
WGBS-seq
RPKM
0.9
0.7
0.5
0.3
0.1
methylation (%)
90
70
50
30
10
104,462,000
104,464,000
104,466,000
104,468,000
B.
Mus musculus GRCm38 chr1:138820000-138855000 (35 kbp)
Lxh9
Lxh9-mRNA
CGI
DEG
Hyper-DMR
+/+
RNA-seq
WGBS-seq
Δ/D329A
RNA-seq
WGBS-seq
RPKM
0.9
0.7
0.5
0.3
0.1
methylation (%)
90
70
50
30
10
138,830,000
138,840,000
138,850,000
C.
Mus musculus GRCm38 chr3:27368000-27379000 (11 kbp)
Ghsr
Ghsr-mRNA
CGI
DEG
Hyper-DMR
+/+
RNA-seq
WGBS-seq
Δ/D329A
RNA-seq
WGBS-seq
RPKM
0.9
0.7
0.5
0.3
0.1
methylation (%)
90
70
50
30
10
27,370,000
27,374,000
27,378,000

Figure 5.13. Genome browser views of regions with altered gene expression and differential methylation. Genome browser views from of the: **A.** *Pou4f1* locus that is de-repressed and exhibits methylation gain over the gene body; **B.** *Lhx9* locus that is upregulated and exhibits methylation gain over both gene body and promoter region; **C.** *Ghsr* locus that is downregulated and exhibits methylation gain over the promoter region. Each coloured block represents a 500bp with 50bp step. Raw sequencing reads are shown. For RNA-seq the colour indicates direction, where red is a forward read and blue is a reverse; libraries were prepared using opposite-strand specific kit. For methylation, red means a methylated call and blue means an unmethylated call, each point indicating an individual C in CpG context. For the mRNA tracks, red indicates genes transcribed from left to right and blue indicates genes transcribed from right to left.

5.3 Aberrant DNA methylation is present in multiple tissues

In order to understand whether abnormal gain of DNA methylation is specific to hypothalamus, I set out to assess other tissues. The pituitary gland was selected for two reasons; it is involved in the growth regulatory axis in the brain but structurally is a separate region from hypothalamus. In addition to pituitary, liver was selected as a very homogeneous tissue, outside of the brain. PBAT libraries were generated for pituitary and liver. Three replicate libraries from liver tissue were generated per genotype. Elena Ivanova kindly generated two to three replicate libraries per genotype from pituitary tissues. The range of duplication was 4.4%-11.1% and 4.4-18.2% in pituitary and liver PBAT libraries, respectively (Table 8 and Table 9). All libraries passed FastQC checks.

Table 8. Pituitary PBAT library details.

Sample ID	Duplication rate	Uniquely mapped deduplicated reads (million)
+/+_1	6.9%	24.7
+/+_2	6.4%	16.2
+/+_3	9.3%	25
Δ/+_1	11.1%	50.1
Δ/+_2	7.1%	23
Δ/+_3	8.9%	25
+/D329A_1	10.2%	48.6
+/D329A_2	10.2%	30
Δ/D329A_1	4.4%	16.3
Δ/D329A_2	5%	15.1
Δ/D329A_3	6%	15.3

Table 9. Liver PBAT library details.

Sample ID	Duplication rate	Uniquely mapped deduplicated reads (million)
+/+_1	17.2%	40.2
+/+_2	9.3%	38.2
+/+_3	16.9%	35.4
Δ/+_1	5.6%	40.9
Δ/+_2	4.8%	35.2
Δ/+_3	4.6%	39.3
+/D329A_1	5.1%	39.1
+/D329A_2	5.6%	40.9
+/D329A_3	9.1%	34.9
Δ/D329A_1	4.4%	36.7
Δ/D329A_2	9.6%	38.3
Δ/D329A_3	18.2%	37.43

Pituitary PBAT libraries of *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ were sequenced first and had a depth of 15.1-24.7 million reads per sample (Table 8). When heterozygous samples were added, their sequencing depth ranged between 23-50.1 million reads per sample, and therefore were down-sampled during analysis to match the mean number of reads in *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$. Since the coverage was lower than usual, 300-CpG genomic tiles were used to assess DNA methylation levels in pituitary. Liver PBAT libraries were sequenced to the depth of 34.9-40.9 million reads per sample (Table 9). For

simplicity of later comparisons across tissues and different developmental stages, liver DNA methylation analysis showed here was also done using 300 CpG tiles.

Similar to hypothalamus, global DNA methylation levels in both pituitary gland and liver showed barely any difference across the four genotypes (Figure 5.14). The global DNA methylation levels in pituitary were at 70.0%, 68.3%, 71.2% and 70.1% in *Dnmt3a*$^{+/+}$, *Dnmt3a*$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329}$, respectively. In liver, they were at 71.8%, 70.4%, 71.5% and 69.9% in *Dnmt3a*$^{+/+}$, *Dnmt3a*$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329}$, respectively. PCA analysis showed that individual replicates of the same genotype clustered together, but were spatially separated based on the genotype (Figure 5.15). This was true for both liver and pituitary, suggesting that within genotype, global methylomes are very reproducible but there are genotype-associated differences.

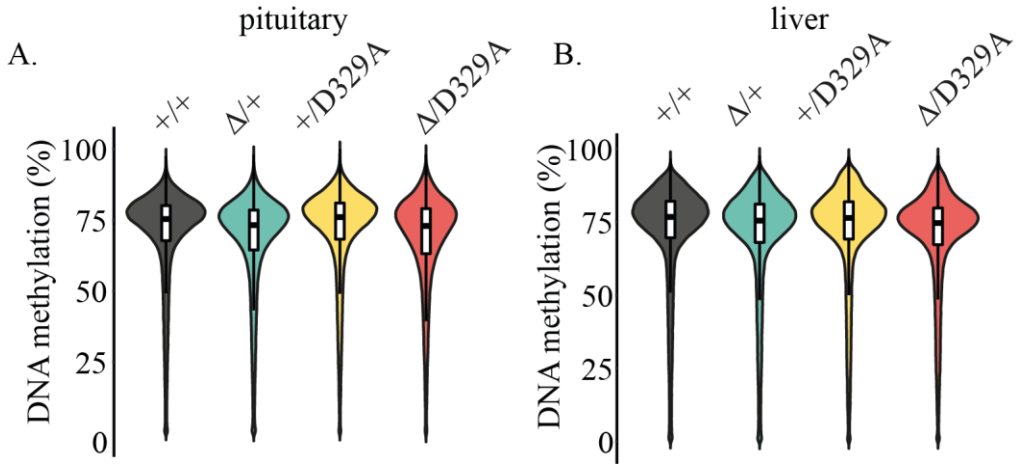


Figure 5.14. Global DNA methylation in pituitary and liver. Beanplots indicating whole genome methylation levels in **A.** pituitary and **B.** liver of adult male mice carrying the alleles shown. Tiles of 300 CpG positions were used. Kruskall-Wallis rank sum test results in p-value <2.2e-16 and is significant for all post-hoc pairwise Wilcoxon rank sum tests, with exception of **A.** pair *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$, where p-value=0.17.

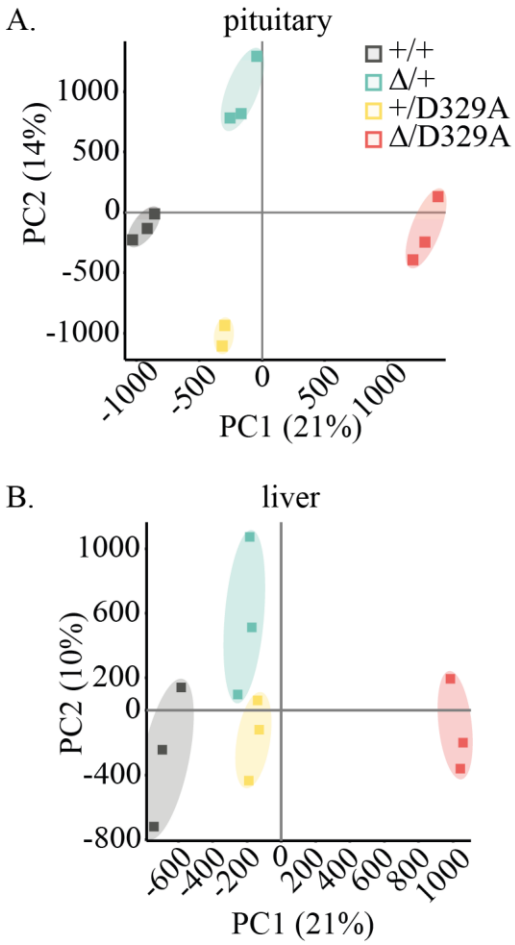


Figure 5.15. PCA plots for global DNA methylation in pituitary and liver. Colour denotes genotype, alleles of *Dnmt3a* are listed. Tiles of 300 CpGs were used, each dot represents an individual mouse.

As in previous analysis, the edgeR tool was used to identify tiles showing differential methylation between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ in the pituitary and liver. In total, 696 genomic tiles were found to be hypermethylated and 323 genomic tiles were hypomethylated in pituitary (Figure 5.16a); meanwhile, 210 tiles were hypermethylated and 78 tiles were hypomethylated in liver (Figure 5.16b). Comparison of *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329}$ in pituitary found 878 hypermethylated DMRs and only 43 hypomethylated DMRs (Figure 5.17c). In liver samples, the same comparison yielded 312 hypermethylated DMRs and only 5 hypomethylated DMRs (Figure 5.16d). These findings mirrored those in hypothalamus, and indicated that an aberrant gain of DNA methylation, caused by presence of the D329A mutation, is also found in other tissues.

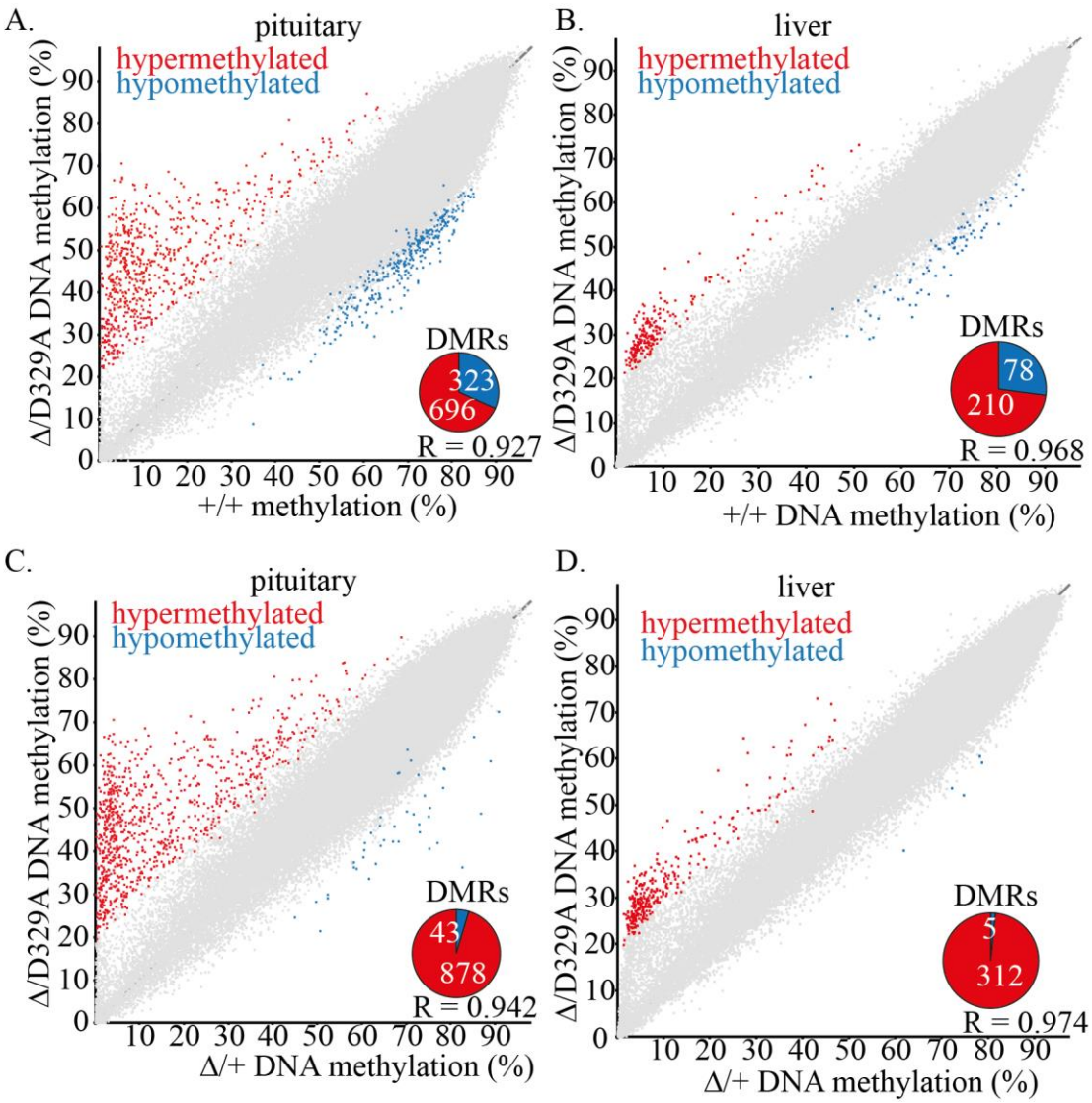


Figure 5.16. DNA methylation differences in liver and pituitary tissues. Scatterplots showing correlation between methylation levels of individual 300-CpG tiles between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ in adult male **A.** pituitary and **B.** liver, and between *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$ in **C.** pituitary and **D.** liver. Differentially methylated tiles were determined using the EdgeR proportion statistic in SeqMonk (p<0.01 corrected for multiple comparisons using Benjamini-Hochberg, methylation difference ≥20%). Pie-chart insets indicate how many of DMRs are hyper- or hypo- methylated.

Out of 600 hypomethylated DMRs originally identified in the hypothalamus, 47 were also found in pituitary, 16 were also found in liver, and only 6 were found in all three tissues (Figure 5.17). In contrast, less than half of hypermethylated DMRs were specific to

hypothalamus. Out of 3,586 hyper-DMRs, 1364 were shared with pituitary, 94 were shared with liver, and 534 were shared between all three tissues.

These findings indicate that aberrant methylation gain could be a result of targeted action, rather than random. The overlap of hypermethylated DMRs across three different tissues strongly implied that aberrant gain of DNA methylation could be an early event during embryonic development, prior to tissue differentiation and specification.

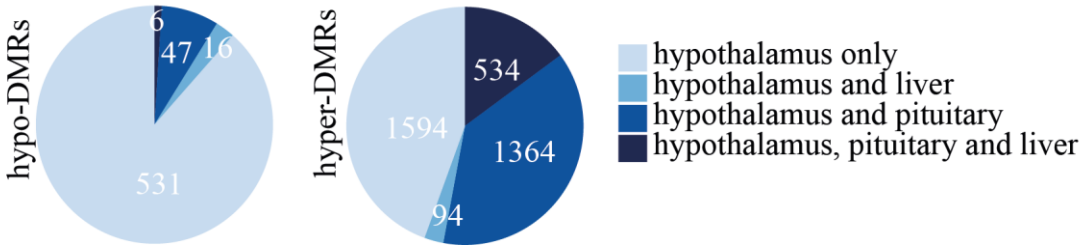


Figure 5.17. Overlap of DMRs in hypothalamus, pituitary and liver. Pie charts showing how many hypo- and hyper- methylated tiles in hypothalamus overlapped the corresponding differentially methylated tiles in pituitary, liver or both.

5.4 DNA methylation gain is progressive in postnatal development

After fertilisation, DNA methylation patterns of terminally differentiated oocyte and sperm cells are largely erased, and are followed by a major phase of epigenetic programming, including DNA methylation, at the period of E4.5-E6.5 (Figure 1.2) (Smith et al., 2012). This is associated with gain of multipotent capacity, required to build different tissues during development. To address the possibility that aberrant methylation was established during this phase of embryonic remethylation, E7.5 mouse epiblasts (Epi) were kindly collected by Jessica Elmer and Wendy Dean from embryos of all four genotypes. E7.5 should correspond to a developmental stage by which the genome is globally remethylated. Four replicates per genotype, each representing a single epiblast, were used to generate PBAT libraries. Duplication rates in these libraries were in the range of 24.8-88.4% and they had sequencing depth of 4.8-27.6 million uniquely mapped reads (Table 10). All libraries passed FastQC checks.

Table 10. E7.5 epiblast PBAT library details.

Sample ID	Duplication rate	Uniquely mapped reads (million)
+/+_1	68.1%	11.9
+/+_2	36.6%	24.9
+/+_3	48.9%	19.5
+/+_4	25.7%	25
Δ/+_1	41.4%	20.5
Δ/+_2	58.5%	15.4
Δ/+_3	49%	16.9
Δ/+_4	41%	21.4
+/D329A_1	75%	7.5
+/D329A_2	35.5%	18.6
+/D329A_3	61.3%	16.8
+/D329A_4	88.4%	4.8
Δ/D329A_1	24.8%	27.6
Δ/D329A_2	28%	24.5
Δ/D329A_3	55.5%	20
Δ/D329A_4	63.7%	5.5

Individual embryos showed comparable but rather variable levels of global DNA methylation, mean level per group was at 77.7%, 77.7%, 74.6% and 74.7%, in *Dnmt3a*$^{+/+}$,

Dnmt3a$^{\Delta/+}$, *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329}$ epiblasts, respectively. Global methylation was more variable in *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$ embryos than in *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/+}$ (Figure 5.18). PCA showed no separation of samples based on the genotype, with exception of two, which corresponded to samples showing lowest global DNA methylation within D329A carriers (Figure 5.19a). Pearson distance and pairwise correlation assessment showed that the samples were correlating quite closely and did not separate based on the genotype (Figure 5.19b, Figure 5.19c). The only outliers, again, were those with the D329A mutation and low global DNA methylation. This analysis suggested that the variation observed is stochastic rather than specific and could be attributable to developmental timing. It is possible that individuals with *Dnmt3a*^{D329A} alleles had a delay in methylation completion in comparison to the *wild-type*.

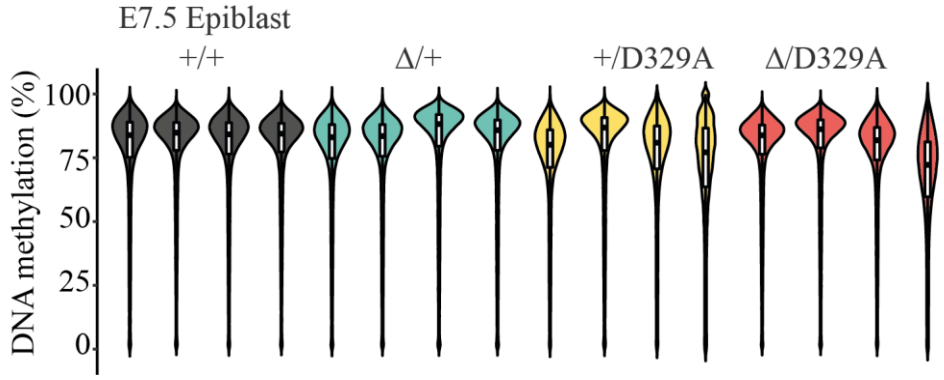


Figure 5.18. Global DNA methylation in E7.5 epiblast. Beanplots indicating whole genome methylation levels assessed over in individual E7.5 epiblasts of different genotypes. Tiles of 300CpG positions were used. Kruskall-Wallis rank sum test using mean tile values for each genotype results in p-value <2.2e-16 and is significant for all post-hoc pairwise Wilcoxon rank sum tests.

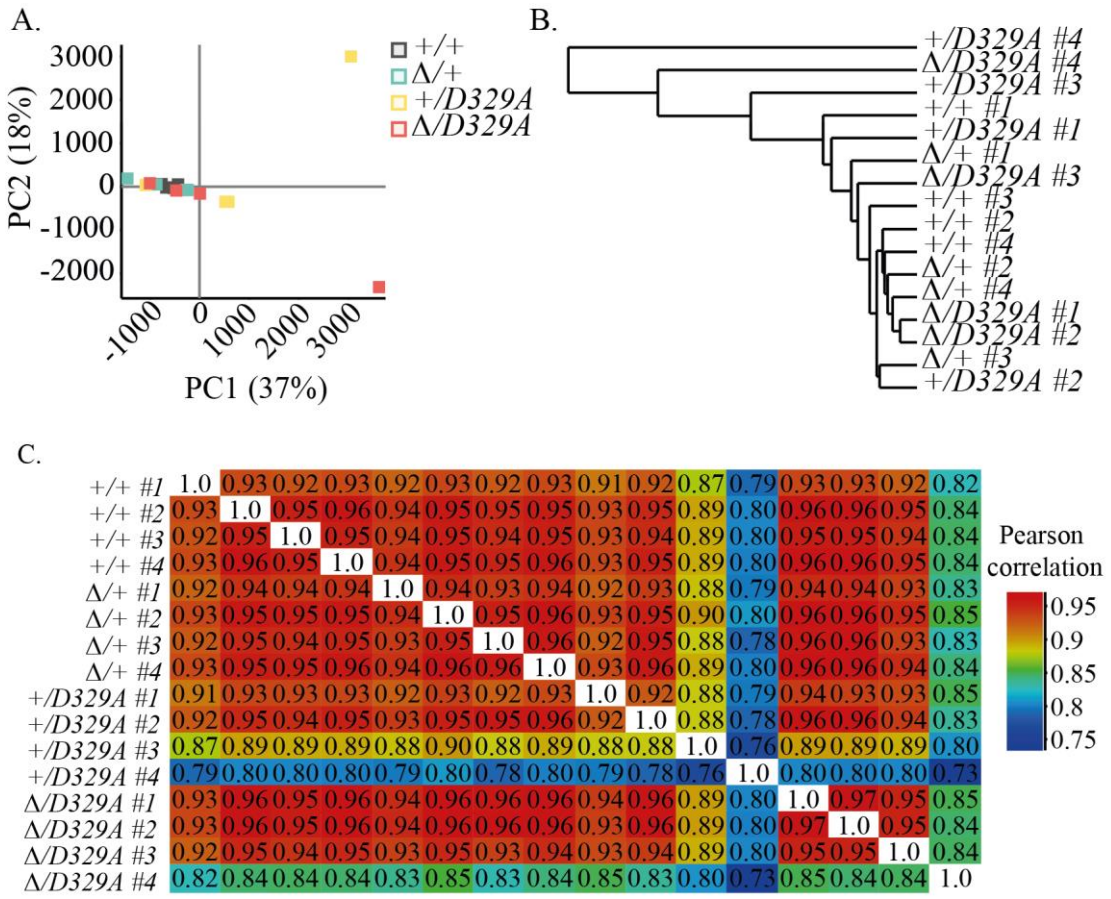


Figure 5.19. E7.5 DNA methylation similarity across samples. A. A PCA plot showing clustering of the majority of the individual E7.5 epiblasts together. **B.** A data similarity tree indicating Pearson correlation distances between individual samples. **C.** A matrix showing pairwise Pearson correlation values for individual sample pairs.

On average, DNA methylation levels over genomic features, such as promoters, gene bodies and intergenic regions, were lower in *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$ embryos, compared to *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/+}$ (Figure 5.20). Interestingly, at CGI promoters DNA methylation levels were also lower in *Dnmt3a*$^{\Delta/+}$ samples. The edgeR tool was used to identify genomic regions sensitive to *wild-type* allele dosage at E7.5 and identified 136 DMRs (0.13% of the genome). All but two DMRs overlapped a CGI, suggesting that at this embryonic stage the susceptibility to dosage is rather specific to areas of high CpG density. Meanwhile, comparison of *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ embryos found only 5 DMRs (0.005% of the genome), all of which were hypomethylated. These findings suggested that the aberrant DNA methylation observed in multiple adult tissues was not

established during the major epigenetic programming event in early embryonic development.

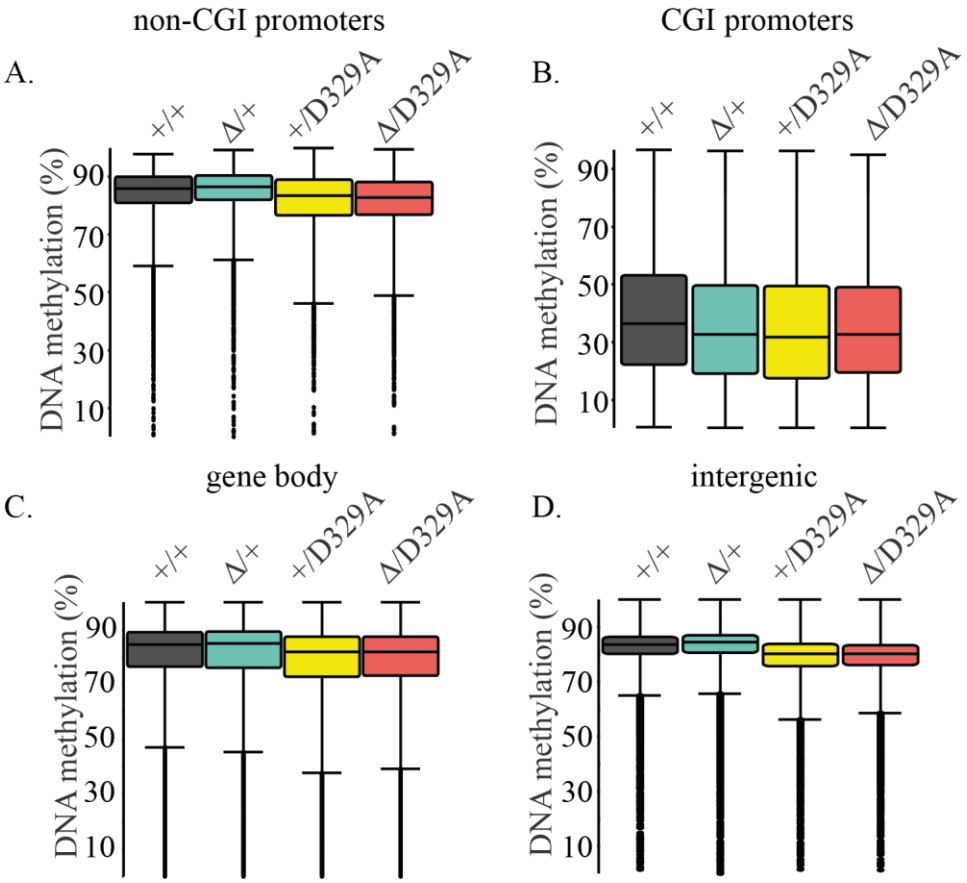


Figure 5.20. DNA methylation in E7.5 embryos over genomic features. DNA methylation levels in E7.5 epiblast at **A.** non-CGI promoters, **B.** CGI promoters, **D.** gene bodies, and **D.** intergenic regions. Kruskall-Wallis rank sum test results in p-value <2.2e-16 and is significant for all post-hoc pairwise Wilcoxon rank sum tests. The exceptions are pairs **B.** *Dnmt3a*$^{\Delta/+}$ and *Dnmt3a*$^{\Delta/D329A}$, where p=0.98, **C.** *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/+}$, and *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$, where p=0.38 and p=0.55, respectively, **D.** *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$, where p= 0.068.

Hypothalamic neurons undergo further epigenetic modelling during postnatal maturation (Li et al., 2014), and so it is possible that aberrant gain of methylation occurred later during brain development. To understand whether hypermethylated domains are a result of brain-specific epigenetic events, hypothalami of *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ mice were collected at postnatal day 1 (P1) and day 25 (P25), and PBAT libraries generated from individual hypothalami. Since the intention was to check the levels of DNA

methylation over hypermethylated domains, only two replicates for each *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ genotype were collected. Libraries had duplication levels ranging from 1.8-2.8% and 17.1-54.2 million uniquely mapped reads, they all passed FastQC checks.

Table 11. P1 and P25 hypothalami PBAT library details. P1 – postnatal day 1, P25 – postnatal day 25.

Sample ID	**Duplication rate**	**Uniquely mapped reads (million)**
P1_+/+_1	1.8%	18.9
P1_+/+_2	2.4%	54.2
P1_Δ/D329A_1	2.4%	35.5
P1_Δ/D329A_2	1.9%	17.1
P25_+/+_1	2.50%	36.3
P25_+/+_2	2.8%	41.3
P25_Δ/D329A_1	3%	51.9
P25_Δ/D329A_2	2%	21.2

Together with E7.5 and adult data, I focussed on domains that are destined to be hypermethylated in *Dnmt3a*$^{\Delta/D329A}$ adult mice. *Dnmt3a*$^{\Delta/D329A}$ E7.5 epiblast was indistinguishable from *Dnmt3a*$^{+/+}$ (Figure 5.21 and Figure 5.22). Meanwhile, at P1, a gain of DNA methylation could be observed at hyper-DMR regions in *Dnmt3a*$^{\Delta/D329A}$ samples. At P25 the methylation increase in hyper-DMRs of *Dnmt3a*$^{\Delta/D329A}$ samples was almost as high as observed in adult hypothalamus. The gain appeared to be evenly distributed over larger domains spanning multiple DMRs (Figure 5.23). These data suggest that DNA methylation gains are progressive throughout development. Furthermore, they are likely

to be established in all tissues independently, which would also explain why not all hypermethylated tiles are shared by hypothalamus, pituitary and liver.

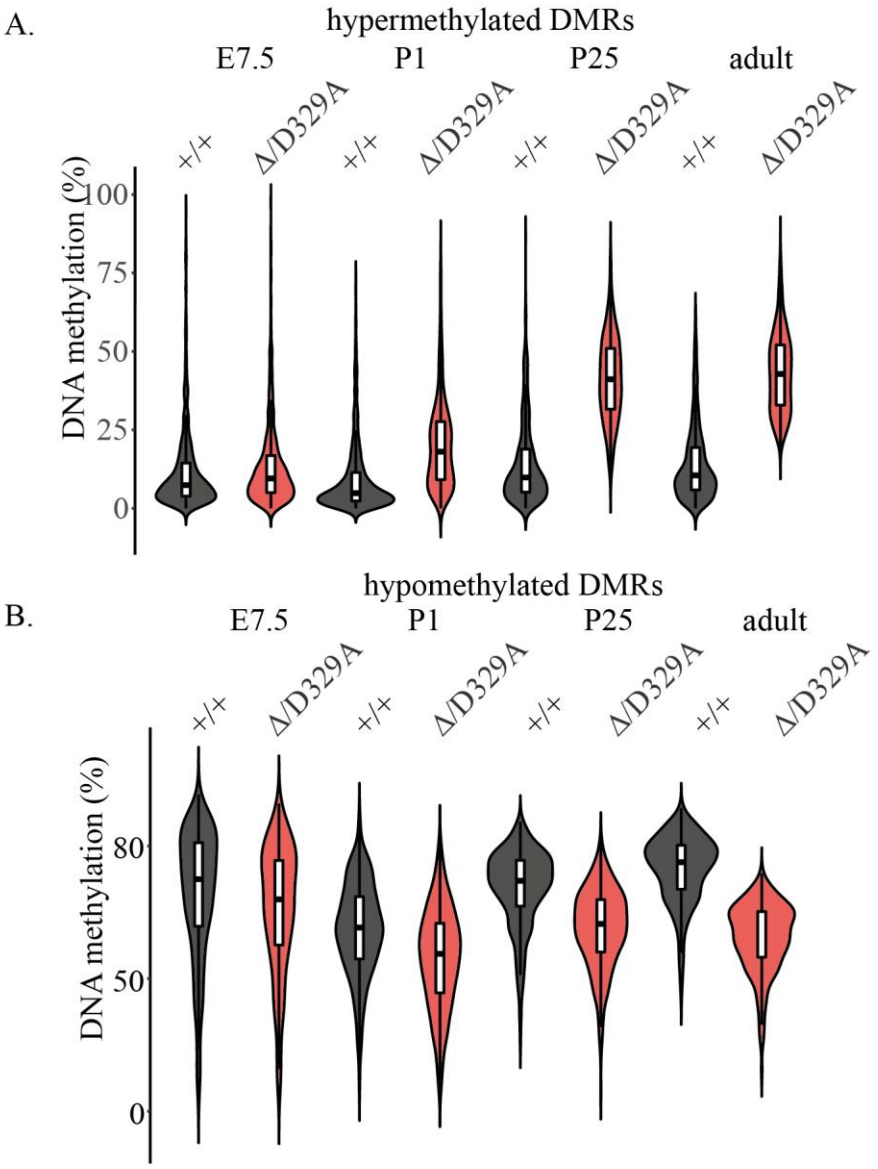


Figure 5.21. DNA methylation changes during development. Beanplots indicating DNA methylation levels over **A.** hyper- and **B.** hypo- methylated DMRs across development in *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ animals. Tiles were quantified over adult hypothalamus DMRs, which were merged if the distance was less than 1kb.

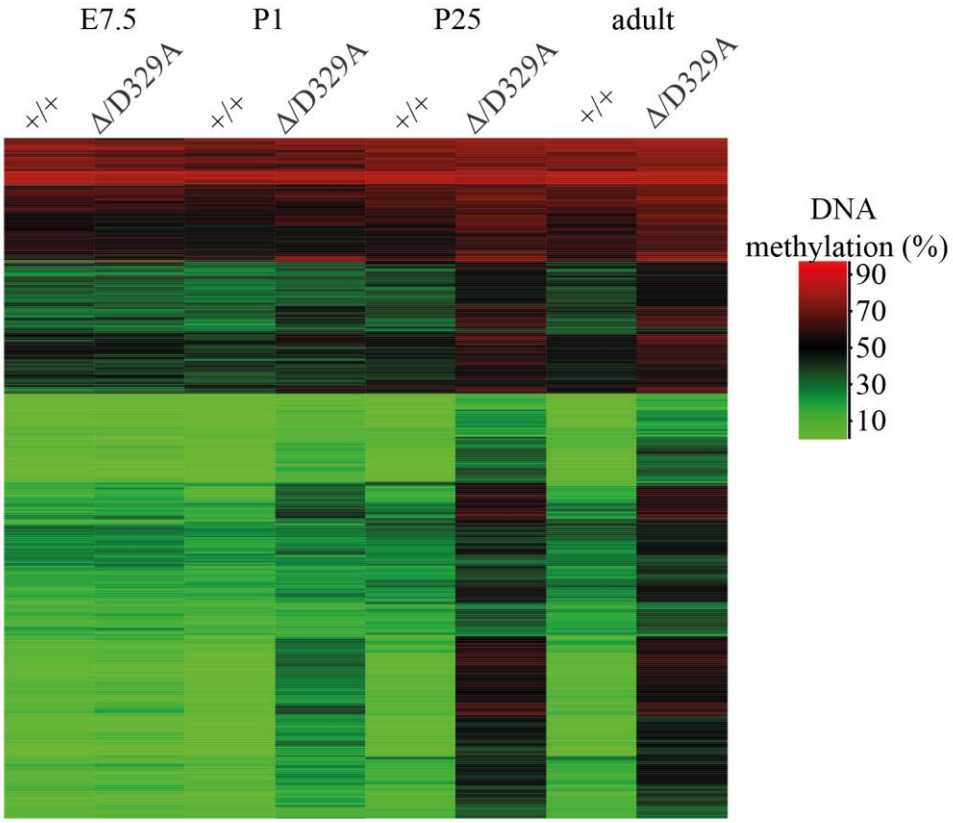


Figure 5.22. DMR heatmap during development. Heatmap showing methylation gain at DMRs over time. Shown are 1915 300-CpG tiles, overlapping DMRs identified between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ in adult (14-week) hypothalamus. DMRs are clustered based on Euclidean method, on the basis of smallest absolute difference between quantitation values.

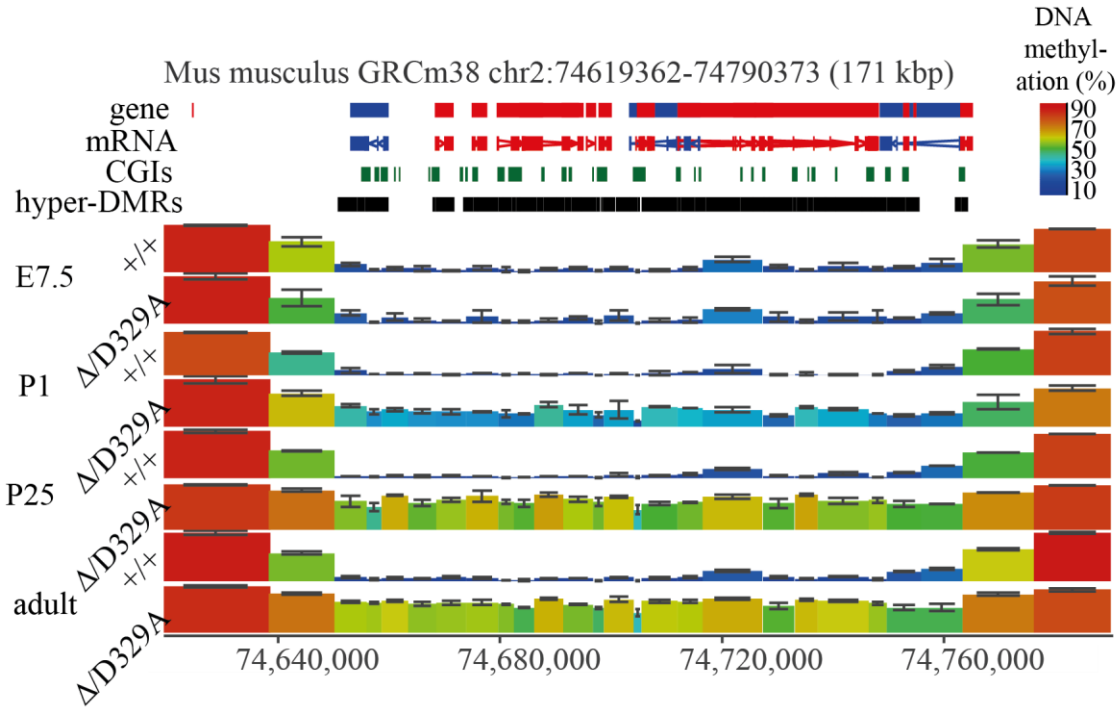


Figure 5.23. Genome browser view of progressive DNA gain during development. Genome browser view of the *Hoxd* gene cluster indicating progressive gain across the whole domain. For gene and mRNA tracks, the colour indicates direction, where red is a forward strand and blue is a reverse strand. Colour-coded blocks indicate tiles of 300-CpG positions. Error bars indicate standard deviation.

5.5 Hypermethylated loci are enriched in developmental transcription factor genes

In order to understand why and how aberrant methylation is established by DNMT3A^{D329A} in the genome, I aimed to identify the unifying feature of these genomic regions. Previously, it was observed that many hypermethylated DMRs fell within large normally unmethylated genomic regions (Figure 5.24), and CGI promoters had a higher DNA methylation in animals carrying the *Dnmt3a*^{D329A} allele. To further investigate this, DMR tiles were separated based on the genomic features, namely promoter, gene body, and intergenic regions, they overlapped. The features were additionally split into CGI-containing or not. Since there were 3584 hypermethylated DMRs and 600 hypomethylated DMRs, a randomly sampled list of 2000 non-DMR tiles was created as a representative of the genome. Surprisingly, I found that over 94.8% of hypermethylated domains overlapped CGI containing promoters, gene bodies, or intergenic regions, while

hypomethylated domains and random tiles overlapped CGI containing features only in 20.3% and 7.25% of cases, respectively (Figure 5.25).

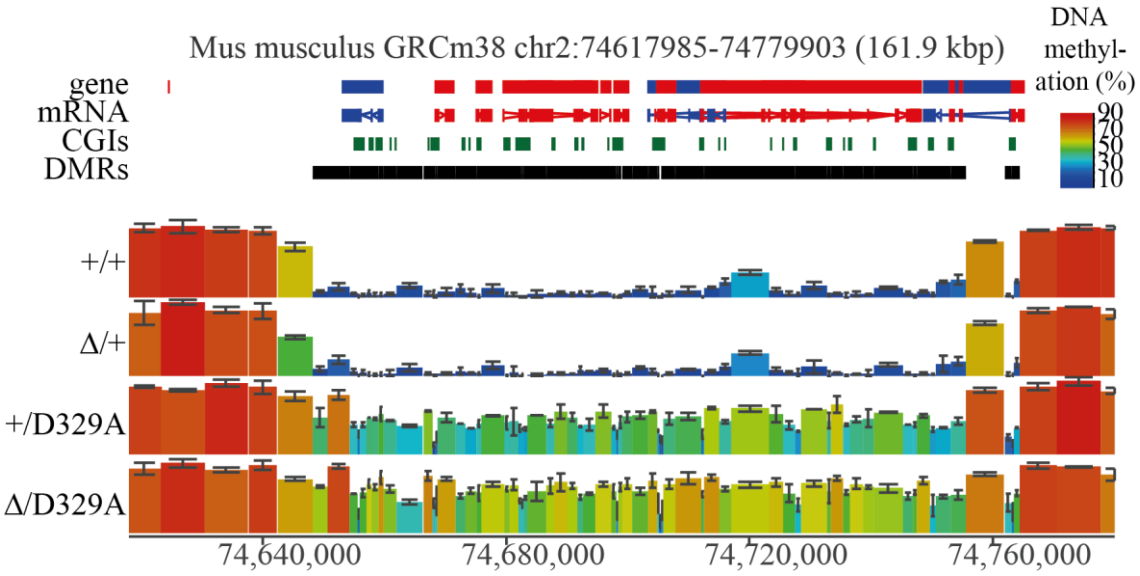


Figure 5.24. *Hoxd* gene cluster showing gain of DNA methylation. Representative genome browser region showing the *Hoxd* hypermethylated domain. For gene and mRNA tracks, the colour indicates direction, where red is a forward strand and blue is a reverse strand. Colour-coded blocks are tiles of 100-CpG positions. Error bars indicate standard deviation.

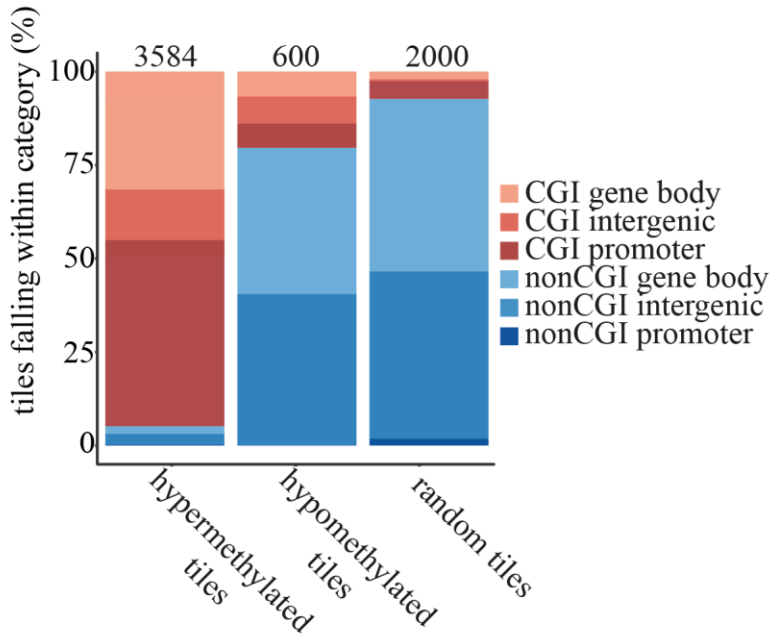


Figure 5.25. DMR separation based on underlying features. Percentages of hypermethylated and hypomethylated tiles (100-CpG) falling within the genomic features indicated, compared with a random set of 100-CpG tiles.

Given that 1) hyper-DMRs were enriched in CGIs, 2) a few developmental genes were abnormally methylated and had altered expression in *Dnmt3a*$^{\Delta/D329A}$ hypothalamus, and 3) hyper-DMR regions show very little DNA methylation in *wild-type* samples, the findings suggested that abnormal DNA methylation might be targeted to the so-called methylation 'valleys' or 'canyons'. Methylation canyons, highly conserved amongst vertebrates, are described as large, normally unmethylated, genomic regions, which harbour key developmental genes and are marked by H3K27me3, a histone modification established by the Polycomb complex (Jeong et al., 2014; Li et al., 2018).

DNA methylation canyons were initially described in hematopoietic cells (Jeong et al., 2014). To test the hypothesis that these domains are gaining aberrant DNA methylation in hypothalamus, I first looked at the DNA methylation valley annotation in hematopoietic stem cells. Out of the 1104 canyons originally defined, 779 were also unmethylated (<10% methylation) in *Dnmt3a*$^{+/+}$ adult mouse hypothalamus. Comparison of DNA methylation levels across canyons in *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ hypothalamus showed a dramatic gain of DNA methylation in the presence of the D329A mutation (Figure 5.26) and suggested these regions might be the target of mutant *Dnmt3a*.

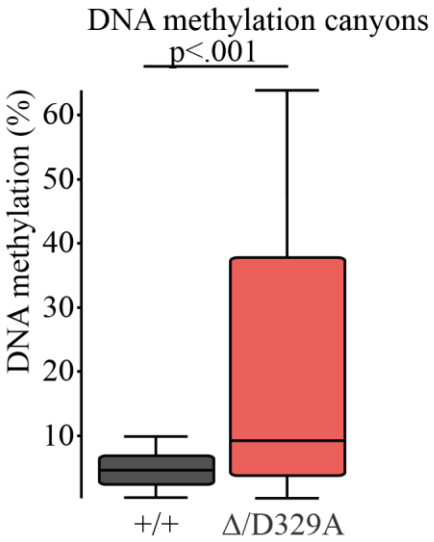


Figure 5.26. DNA methylation levels over DNA methylation valleys. Distribution of methylation values over DNA methylation canyon probes in *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ mice. n = 779, Mann-Whitney Test (p <0.01, U_A= 432773, z = -14.57).

GO analysis was used to find gene groups overrepresented within hypo- and hyper-DMRs in *Dnmt3a*$^{\Delta/D329A}$ animals. I found that hyper-DMRs were highly significantly enriched in genes encoding transcription factor/regulator activity, DNA binding, morphogenesis and differentiation genes. Hypo-DMRs were also enriched for morphogenesis and transcription factor/regulator activity genes, but to a lower significance (Figure 5.27). GO analysis of genes showing hypermethylation in hypothalamus, pituitary and liver showed the same GO terms being enriched, suggesting the same category of genomic loci is targeted in the different tissues as well (Figure 5.28). This further supported the prediction that differential methylation is targeted to DNA methylation valleys.

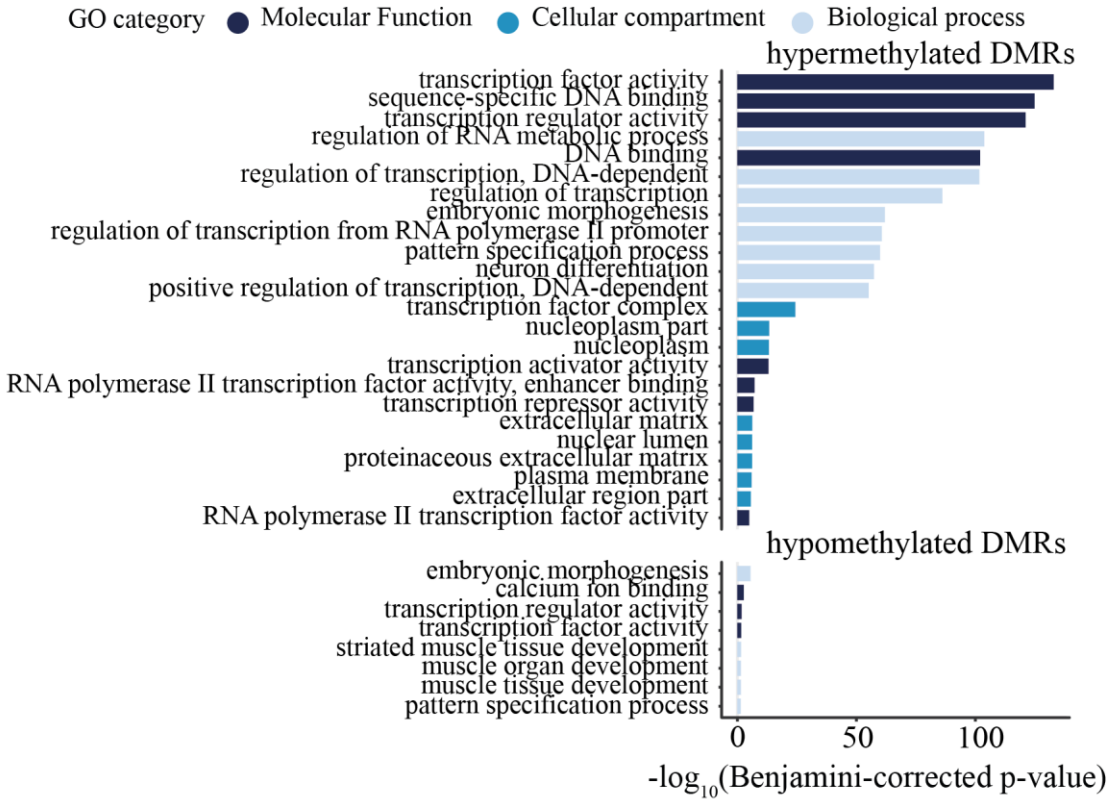


Figure 5.27. GO analysis of hypothalamus DMRs. GO analysis of genes overlapping differentially methylated tiles. Unless there were less significant terms, the top 24 most significant terms per category are shown. Benjamini-Hochberg corrected *p*-value <0.01 was used as a threshold.

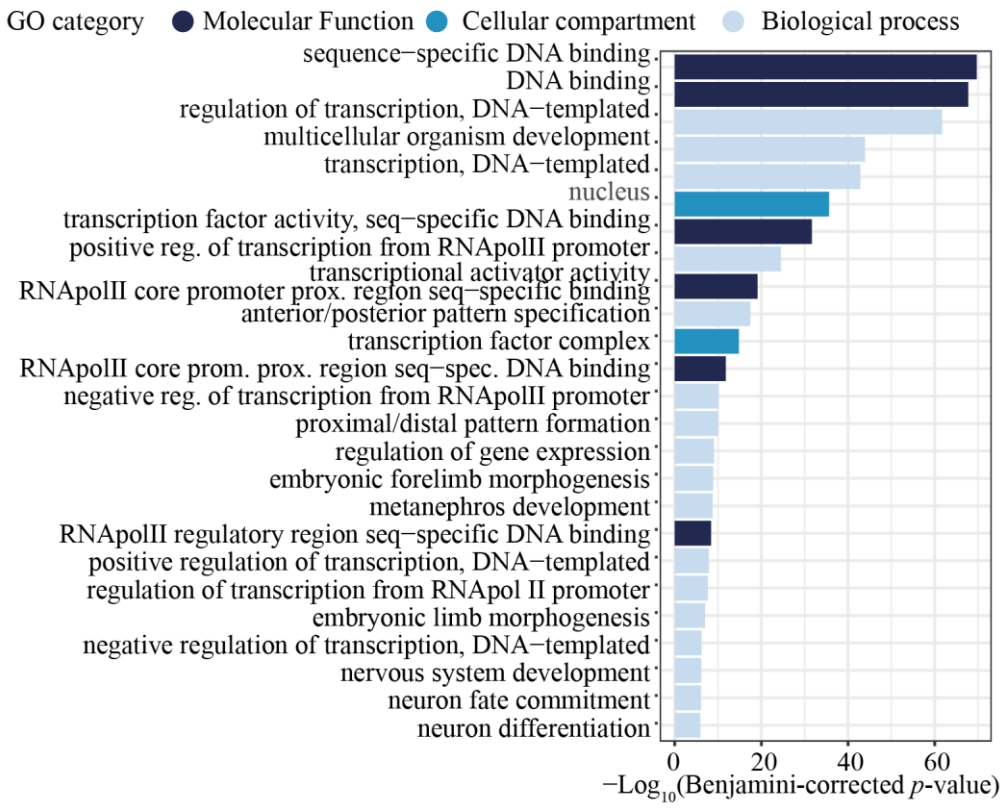


Figure 5.28. GO analysis of hypermethylated DMRs shared between different tissues. Only genes overlapping hypermethylated tiles in hypothalamus, pituitary and liver were included. Top 25 most significant terms per category are shown. Benjamini-Hochberg corrected *p*-value <0.01 was used as a threshold.

5.6 Non-CpG methylation changes as evidence for altered DNMT3A^{D329A} recruitment

Non-CpG (CpH) methylation becomes particularly abundant in neural tissues, where it is established postnatally by DNMT3A (Lister et al., 2013; Varley et al., 2013; Xie et al., 2012). Importantly, CpH methylation is not maintained at DNA replication, so its presence is a record of *de novo* methylation activity since the last replication event and is believed to be a signature of where DNMT3A was bound (Ramsahoye et al., 2000). Thus, using the PBAT datasets, I investigated CpH methylation in adult hypothalamus as a read-out of DNMT3A localisation and activity. Because of the abundance of CpH sites in the genome, I restricted the analysis to two representative autosomes, chromosomes 2 and 11. CpH methylation levels showed strong evidence for haploinsufficiency at the molecular level. Global CpH methylation was 1.8% in *Dnmt3a*$^{+/+}$ and 1.2% in *Dnmt3a*$^{\Delta/+}$

(Figure 5.29a). Similarly, levels of CpH methylation in *Dnmt3a*$^{+/D329A}$ were at 1.6% and in *Dnmt3a*$^{\Delta/D329A}$ at 0.9%. The smaller global reduction between *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{+/+}$ and between *Dnmt3a*$^{\Delta/D329A}$ and *Dnmt3a*$^{\Delta/+}$ suggests some decrease in methyltransferase activity by the mutant protein. Strikingly, however, the pattern over hypermethylated DMRs was reversed. In *Dnmt3a*$^{+/+}$ CpH methylation at hypermethylated DMRs was 1.4% and in *Dnmt3a*$^{\Delta/+}$ it was 0.9%. In contrast, levels at *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$ were 1.6% and 1.3%, respectively (Figure 5.29b).

It is important to note that the Imprint DNA modification kit used to convert unmethylated cytosines to uracil has approximately 99% conversion rate. This means that non-CpG methylation levels reported here are close to expected 1% non-conversion rate. While the values should not be considered as a net quantitative read-out, the relative levels of non-conversion should be the same in all samples, and therefore the differences observed between the genotypes is likely to be real. This difference is clearly visible when assessing the samples in the genome browser (Figure 5.30). Comparison of relative CpH methylation across pre- and postnatal development showed that this effect first becomes evident in the P1 hypothalamus (Figure 5.29c). In summary, these data demonstrate that activity of the DNMT3A^{D329A} protein is aberrantly localised preferentially to hypermethylated DMRs in the postnatal brain.

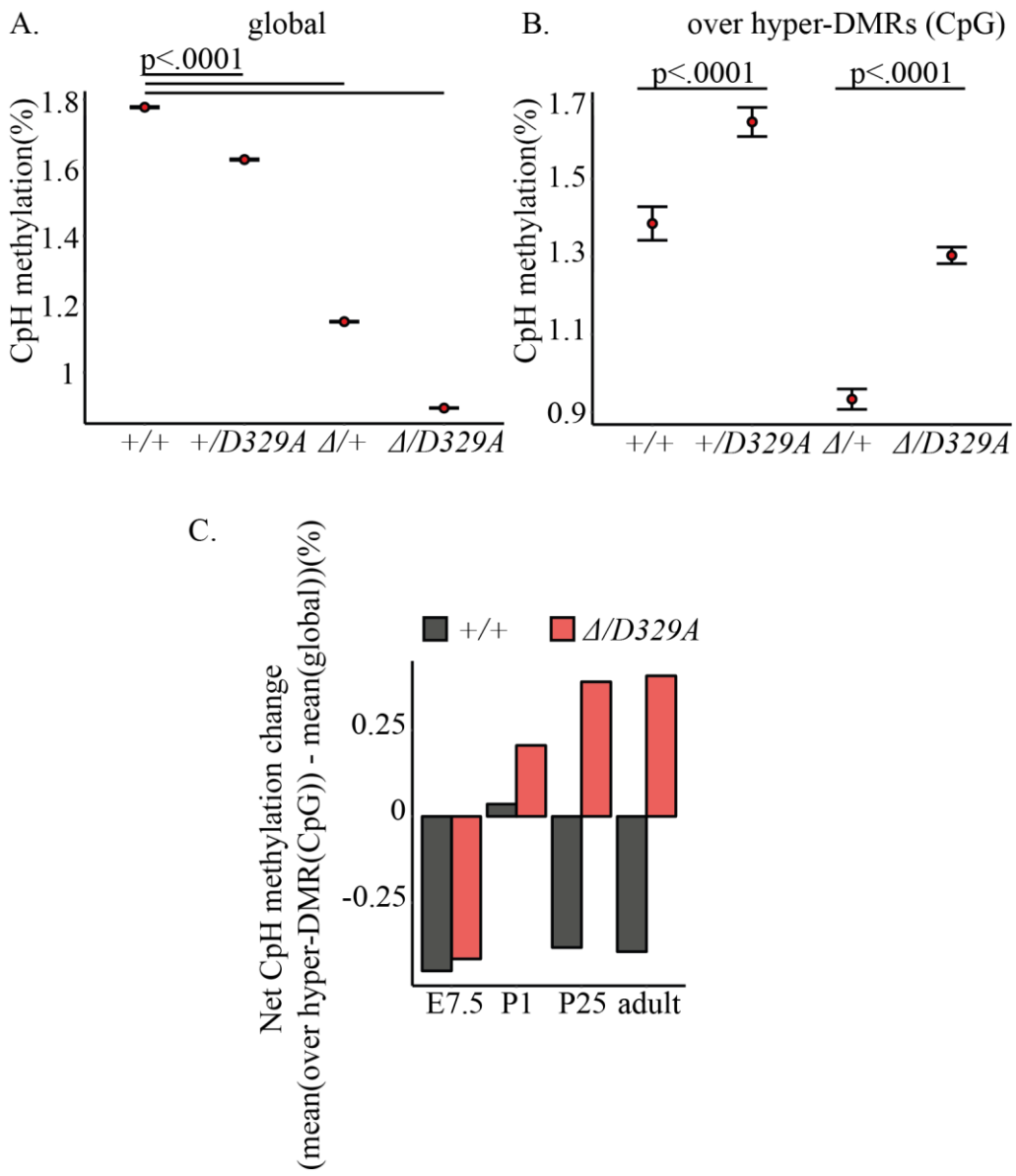


Figure 5.29. CpH DNA methylation. **A.** Global mean CpH methylation values across different genotypes, **B.** Mean CpH methylation values over hypermethylated DMR regions. Error bars indicate standard error of the mean. Pairwise comparisons were done using a two-tailed t-test. **C.** Barplot indicating the net difference in mean CpH methylation between hyper-DMRs and global levels for *Dnmt3a*$^{+/+}$ and *Dnmt3a* $^{\Delta/D329A}$ mice across development (E7.5 epiblast, P1, P25 and adult hypothalamus). 1000-CpH tile quantitation for chromosome 2 and 11 was used as representative.

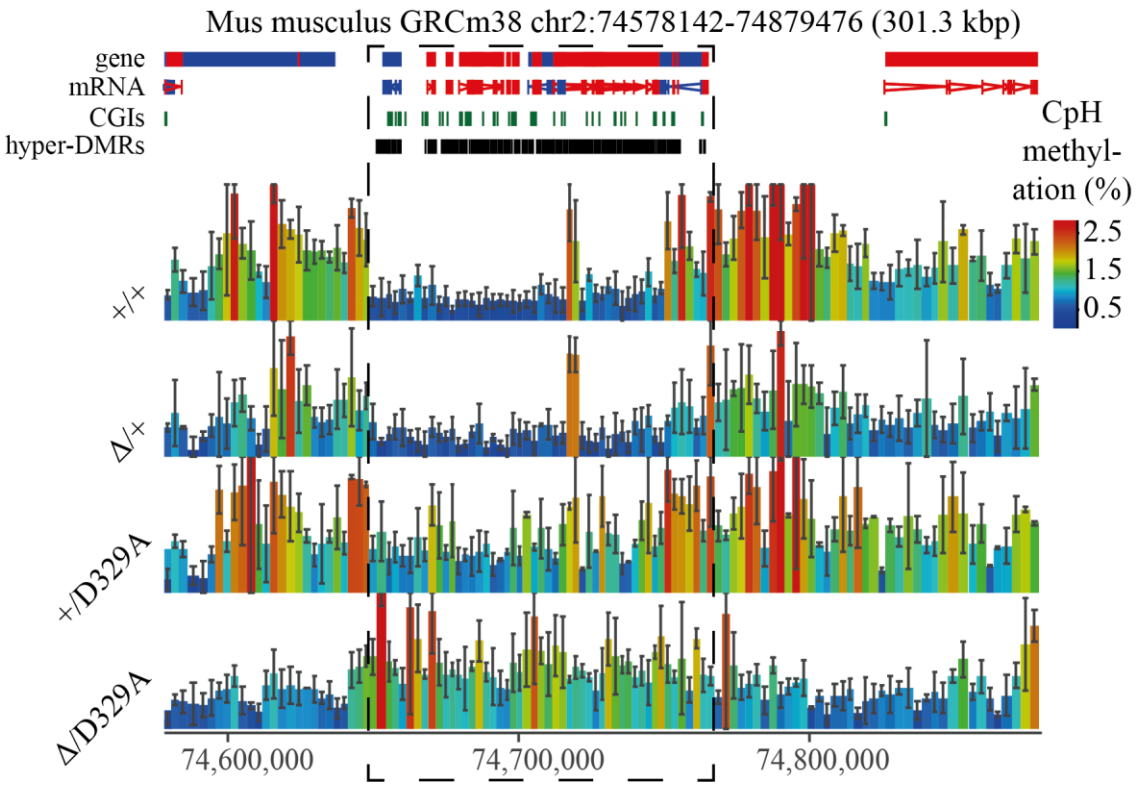


Figure 5.30. Genome browser view of CpH methylation. Representative genome browser view of CpH methylation in adult male hypothalamus. Methylation levels appear to be dependent on the number of alleles present, with a decrease observed in *Dnmt3a*$^{\Delta/+}$ compared to *Dnmt3a*$^{+/+}$. Notably, there is an increase of methylation over hyper-DMR regions (highlighted by the box) in the presence of *Dnmt3a*^{D329A}. Each bar represents a 1000-CpH tile. For gene and mRNA tracks, the colour indicates direction, where red is a forward strand and blue is a reverse strand. Error bars indicate standard deviation.

5.7 Summary and discussion

In this chapter, I show that while most of the global DNA methylation in *Dnmt3a*^{D329A} adult hypothalamus is similar to *wild-type*, approx. 2% of the genome shows aberrant gain of DNA methylation. The targets of this gain are DNA methylation valleys, where many developmental transcription factor genes are located. Some, but not all, of those transcription factors show altered gene expression levels associated with DNA methylation gain. Aberrant DNA methylation can be observed in multiple tissues and occurs progressively throughout development, as a result of differential localisation of DNMT3A to the genome.

The lack of DNA methylation changes over gene bodies appears to agree with findings by Baubec et al. (2015) in ESCs, where they find that it is DNMT3B but not DNMT3A localising to and methylating gene bodies. In *Dnmt3a*^{D329A} E7.5 embryos, I observed a stochastic delay in *de novo* methylation, although it was not localised to particular genomic features. In adult tissues, however, I did not observe a loss of DNA methylation attributable to the D329A mutation and found no evidence that this mutation alters DNA methylation levels at gene bodies. There might be delays in embryonic development, or some genotype-unrelated developmental variation in the litters, meaning they are yet to complete genomic methylation at the time of collection. The number of replicates assessed does not allow to definitively distinguish the cause of lower DNA methylation in some samples. Importantly, these differences do not persist and may be recovered with time, or DNMT3B may be able to compensate during embryogenesis, supported by previous observations that DNMT3A and DNMT3B have partial redundancy in the post-implantation embryo (Dahlet et al., 2020; Greenberg et al., 2017; Okano et al., 1999).

The finding that the D329A mutation results in aberrant hypermethylation was somewhat unexpected, given the original hypothesis that this mutation is a loss-of-function. Nonetheless, the growth retardation found in *Dnmt3a*$^{+/D329A}$ *and Dnmt3a*$^{\Delta/D329A}$, but not in *Dnmt3a*$^{\Delta/+}$, in Chapter 4 was an indication that it could be a gain-of-function, and DNA methylation patterns further confirm this hypothesis. Similarly to our observation, Heyn et al. (2019) found that human fibroblasts and peripheral blood leukocytes from patients with microcephalic dwarfism carrying an orthologous DNMT3A mutation exhibited gain of DNA methylation over developmental gene loci. Their mouse model with a W326R point mutation in the PWWP domain showed a similar molecular pattern. The fact that hypermethylated domains observed in both studies are rich in CGIs, transcription factor loci and located to DNA methylation valleys suggest that a particular chromatin environment might be recruiting mutant DNMT3A to these locations. Further work on this hypothesis can be found in Chapter 6.

DNA methylation is traditionally considered to be a repressive mark at the promoter region, but here I see that gain of DNA methylation over genes can have multiple outcomes. I showed examples of it associated with decrease in expression and methylation increase over promoter, but also cases where DNA methylation over promoters coincided with increased expression. It is possible that some of these genes are harboured in DNA methylation valleys and are under a different control, such as that of

Polycomb repressive complexes (Li et al., 2018). In such cases, DNA methylation might be disrupting the normal repressive mechanisms in place and allow de-repression. Some of the genes show gain of DNA methylation over gene bodies, and in a similar observation, *Hox* genes were activated when DNA methylation was targeted to gene bodies within these regions in tumours (Su et al., 2018). Here, I found some examples of gain of DNA methylation across a large area but a 'dip' in DNA methylation over the promoter, which could potentially create space for required transcription factor binding to initiate expression. Another important consideration is homogeneity of the assessed tissue. Hypothalamus being a very heterogeneous tissue (Chen et al., 2017), it is unclear whether DNA methylation and transcriptional changes occur in the same cells, and what is the extent of their overlap.

Consistently with the onset of the phenotype, hypermethylation was established postnatally by the DNMT3A^{D329A}. *Dnmt3a* is expressed throughout pre- and postnatal development; in prenatal development, the predominant transcript isoform is *Dnmt3a2*, while in postnatal tissues it is *Dnmt3a1* (Chen et al., 2002; Manzo et al., 2017). The only difference is that the longer isoform contains an N-term intrinsically disordered domain, although there are no studies linking the N-terminal disordered and PWWP domains in DNMT3A to date. Similarly, it might be the changing environment and transition from multipotency to a differentiated state in tissues that is facilitating mutant DNMT3A. Alternatively, *Dnmt3b* is very highly expressed throughout early prenatal development and is responsible for the majority of *de novo* methylation (Auclair et al., 2014). Therefore, in this context, the aberrant localisation of the mutant DNMT3A may be rescued by heterodimerisation with DNMT3B (Li et al., 2007b; Emperle et al., 2014).

In summary, here I show that *Dnmt3a*^{D329A} tissues exhibit progressive aberrant gain of DNA methylation at DNA methylation valleys due to altered DNMT3A^{D329A} localisation. Methylation affects expression of transcriptional regulator genes harboured in these DNA methylation valleys.

6 Mutant DNMT3A targets H3K27me3-marked chromatin

Declaration of author contributions

ChIP-seq datasets were generated together with Courtney Hanna.

An earlier version of this chapter has been published:

Sendžikaitė, G., Hanna, C.W., Stewart-Morgan, K.R., Ivanova, E., Kelsey, G., (2019). A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. Nature Communications, 10, Article 1884. https://doi.org/10.1038/s41467-019-09713-w

In the previous chapter, I found that DNMT3A^{D329A} preferentially localises to and methylates DNA methylation canyons, where many developmental transcription factor genes are located (Jeong et al., 2014; Li et al., 2018). The expression of these genes is controlled by the Polycomb repressive complex and they are marked by both activating H3K4me3 and silencing H3K27me3 marks, referred to as bivalent chromatin. This is thought to allow easy and well-controlled spatiotemporal activation and silencing depending on requirements in different tissues. Bivalent chromatin and DNA methylation tend to be mutually exclusive, and therefore I set out to test the hypothesis that the D329A mutation allows aberrant targeting of DNA methylation into these domains.

6.1 Aberrant DNA methylation is targeted to H3K27me3 domains

I hypothesised that DNMT3A^{D329A} hypermethylated DMRs would be enriched in H3K4me3 and H3K27me3. To assess this, I generated H3K4me3, H3K27me3 and H3K36me3 chromatin immunoprecipitation-sequencing (ChIP-seq) datasets from adult hypothalamus (Appendix 2). Approx. 2.5% of a whole hypothalamus volume was used

for each immunoprecipitation and all pull-downs were done from the same chromatin. Three *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ mice were used per genotype.

To analyse ChIP-seq libraries, 2kb genomic tiles with 1 kb step were used and the reads were quantified correcting for total read count per million reads (RPKM values). First, I looked at the distribution of RPKM values for *Dnmt3a*$^{+/+}$ input samples (Figure 6.1), and excluded tiles outside of the bell-shaped distribution area (i.e. >1.75 RPKM) in input samples. This way, genomic locations that show enrichment due to potential duplicate reads mapping over the same region are eliminated. Assessment of samples using Pearson correlation distance and PCA indicated that samples clustered strongly based on precipitated histone mark but not based on the genotype (Figure 6.2 and Figure 6.3). Assessment of cumulative distribution indicated that individual samples had similar distributions for the same mark, and had a strong enrichment of high value tiles at the top 10% of distribution, while input, as expected, had a more even distribution of values (Figure 6.4).

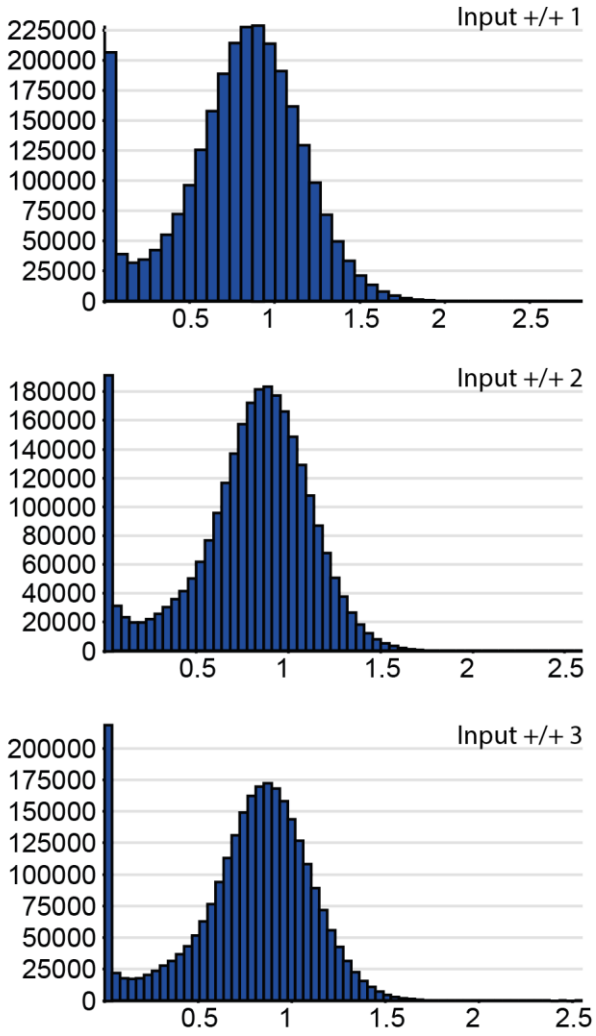


Figure 6.1. ChIP-seq input sample genomic tile distribution. Histograms show the distribution of tile values in *Dnmt3a*$^{+/+}$ input samples.

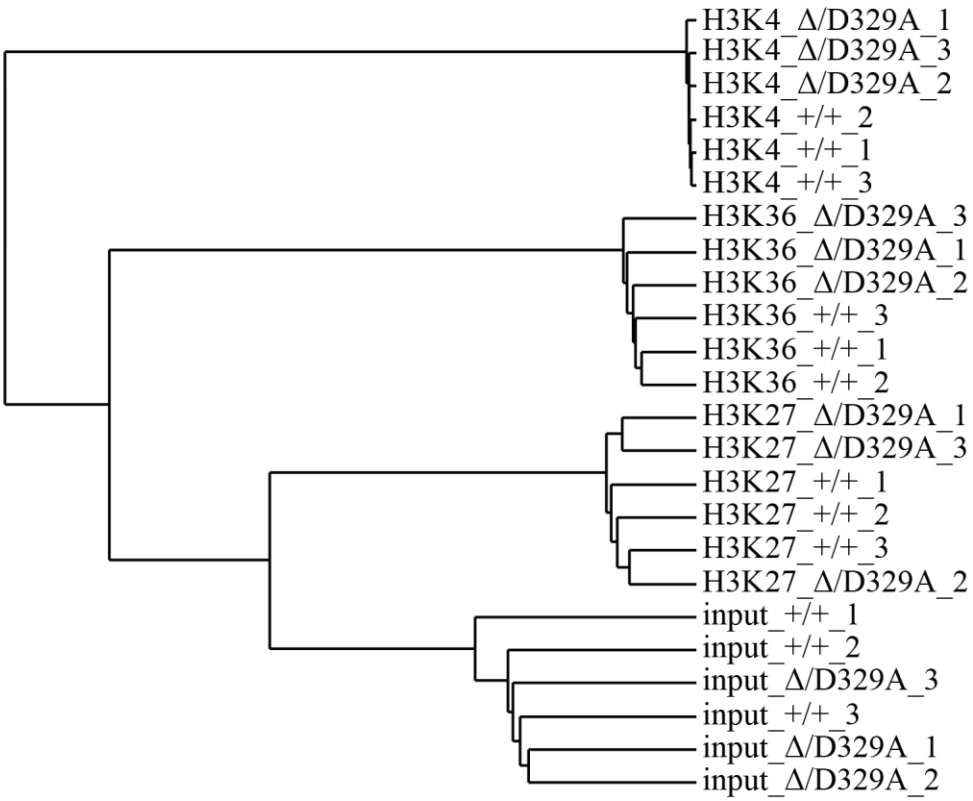


Figure 6.2. ChIP-seq library similarity tree. A data similarity tree indicating Pearson correlation distances between individual ChIP-seq samples, histone PTMs and genotypes are listed.

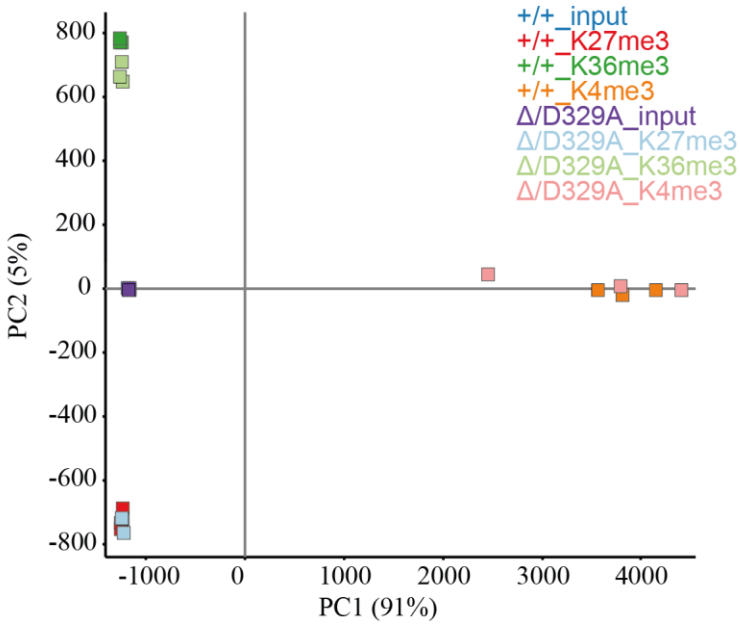


Figure 6.3. PCA of ChIP-seq samples. PCA was done on 2kb tiles, with 1kb step. Each dot represents a replicate sample coming from an individual mouse, colour denotes the genotype.

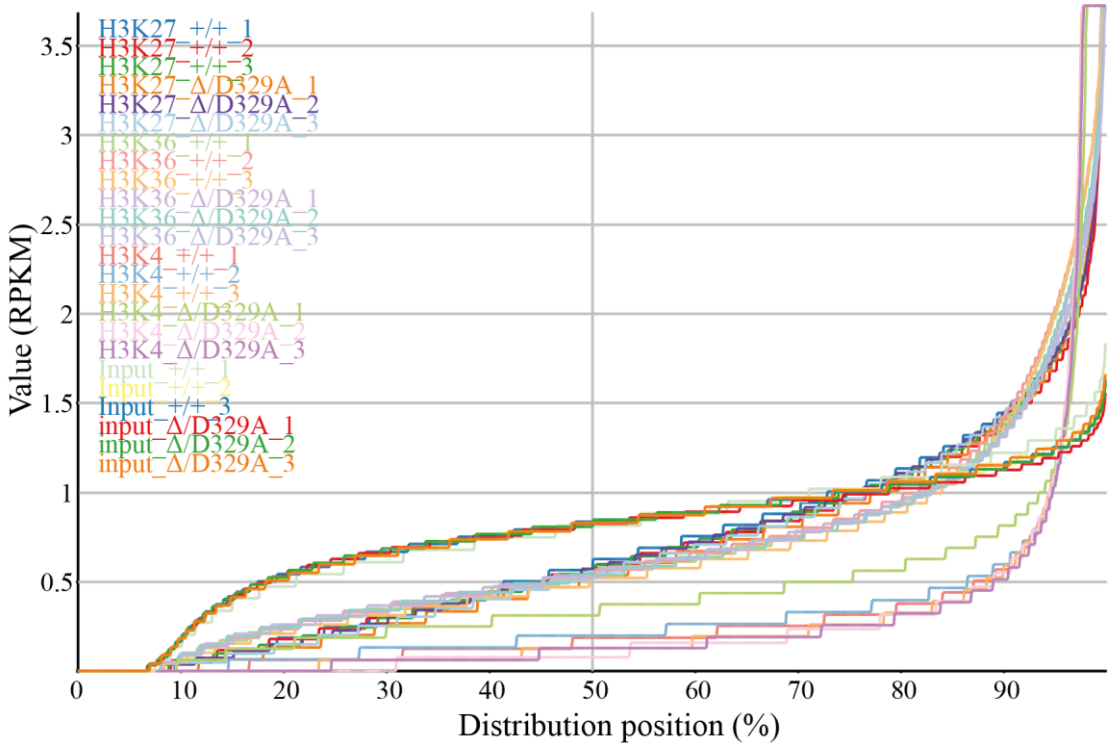


Figure 6.4. Cumulative distribution plot for ChIP-seq datasets. Distribution of genomic tiles in different samples. The tiles are ranked from lowest to highest and indicate values found in different datasets at those distribution fractions.

To see whether histone PTMs exhibited expected localisation in the genome, I assessed the patterns of PTM enrichment over all mRNAs. As predicted, H3K36me3 was enriched over mRNA body, H3K4me3 was strongly enriched at promoter regions, and H3K27me3 was somewhat enriched at promoters but did not show any particular pattern related to mRNA genomic areas (Figure 6.5). Finally, I assessed the genomic landscape on the genome browser. I could clearly see that high H3K4me3 enrichment was located over promoter and TSS area (Figure 6.6). The associated genes had enrichment of H3K36me3 over the gene body, suggesting they are actively expressed in hypothalamus. Some genomic locations showed bivalent trend and had both H3K4me3 and H3K27me3. Bivalent genes with relatively low H3K4me3 levels had a strong enrichment of H3K27me3, suggesting these genes are silenced. In general, initial assessment provided me with the confidence that ChIP-seq datasets are of good quality.

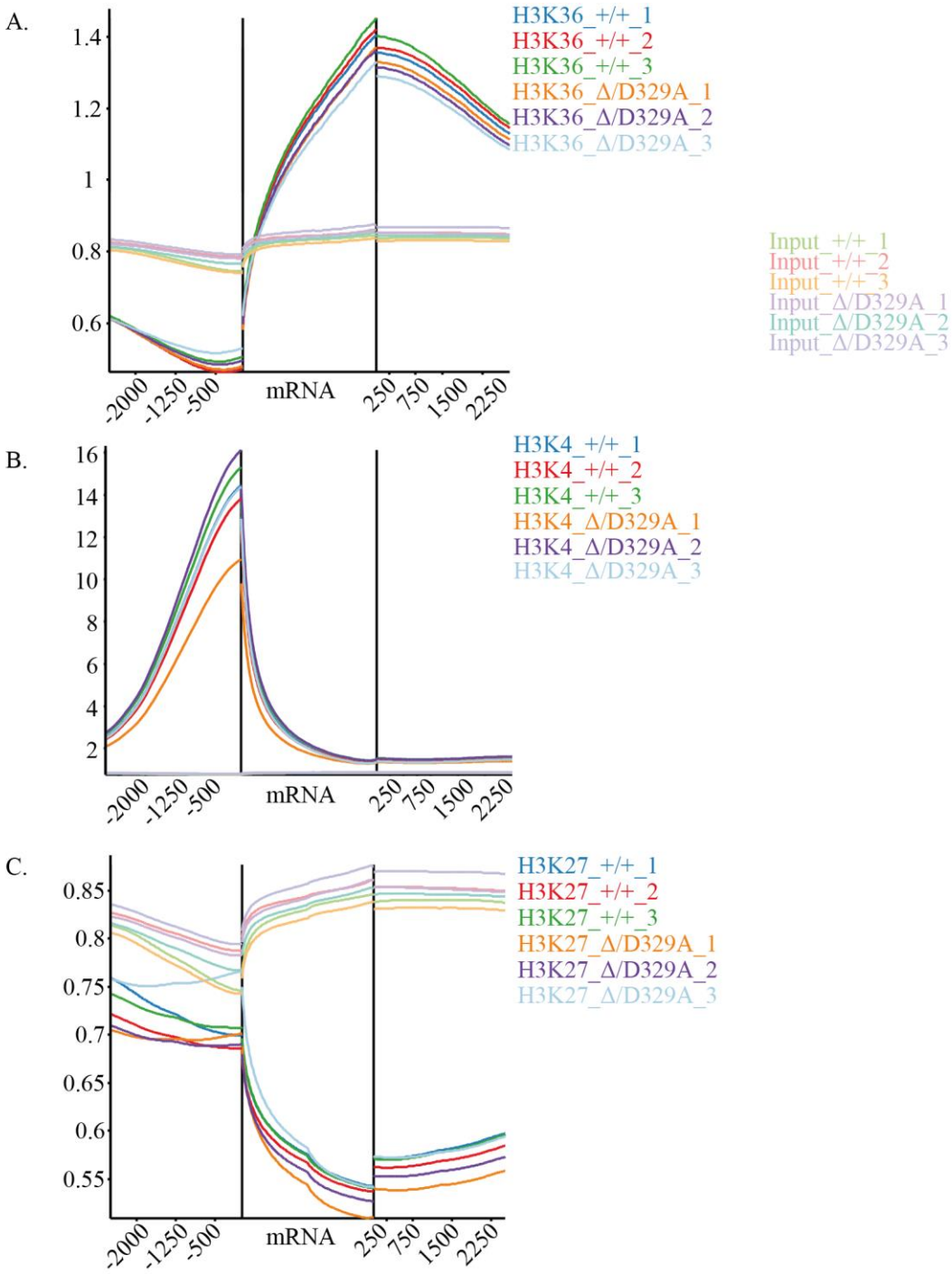


Figure 6.5. Trendplots showing histone PTM library patterns over mRNAs. Aggregated RPKM values for **A.** H3K36me3, **B.** H3K4me3, and **C.** H3K27me3 over all mRNAs ±2,500bp. Input samples are shown in all plots, in section **B.** input lines are indistinguishable from the x axis. Distance over mRNA is relative.

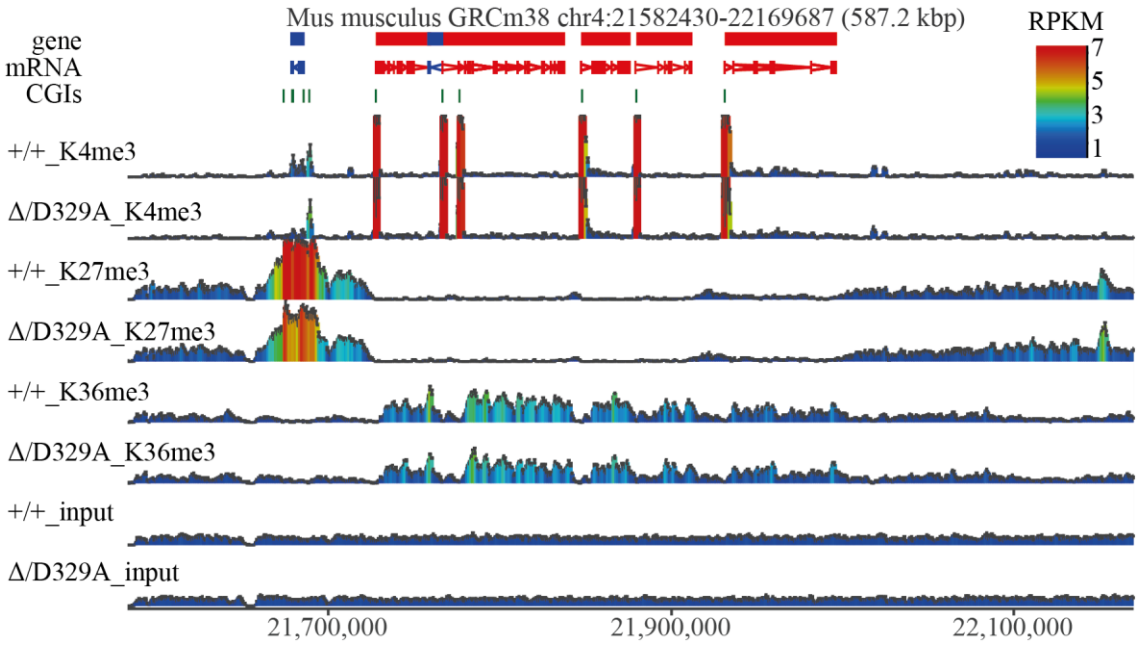


Figure 6.6. Genome browser view of ChIP-seq libraries. Representative genome browser view of adult male hypothalamus H3K4me3, H3K27me3, H3K36me3 and input datasets grouped per genotype. Each colour-coded block represents a 2kb window with 1kb step size. For gene and mRNA tracks, the colour indicates direction, where red is a forward strand and blue is a reverse strand. Error bars indicate standard error of the mean.

Globally, enrichment for H3K4me3, H3K36me3 and H3K27me3 marks was correlated highly between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ samples for all three histone marks, and I did not observe a marked shift in the distribution of these marks in the presence of the *Dnmt3a*^{D329A} allele (Figure 6.7). When the distribution of hypo- and hyper-DMRs over histone modifications was assessed, there was little to no difference between the two sets of DMRs in H3K4me3 and H3K36me3 (Figure 6.8a and Figure 6.8c). However, hypo-DMRs were found to overlap regions with low enrichment of H3K27me3, while hyper-DMRs were overlapping areas with mid- and high H3K27me3 enrichment (Figure 6.8).

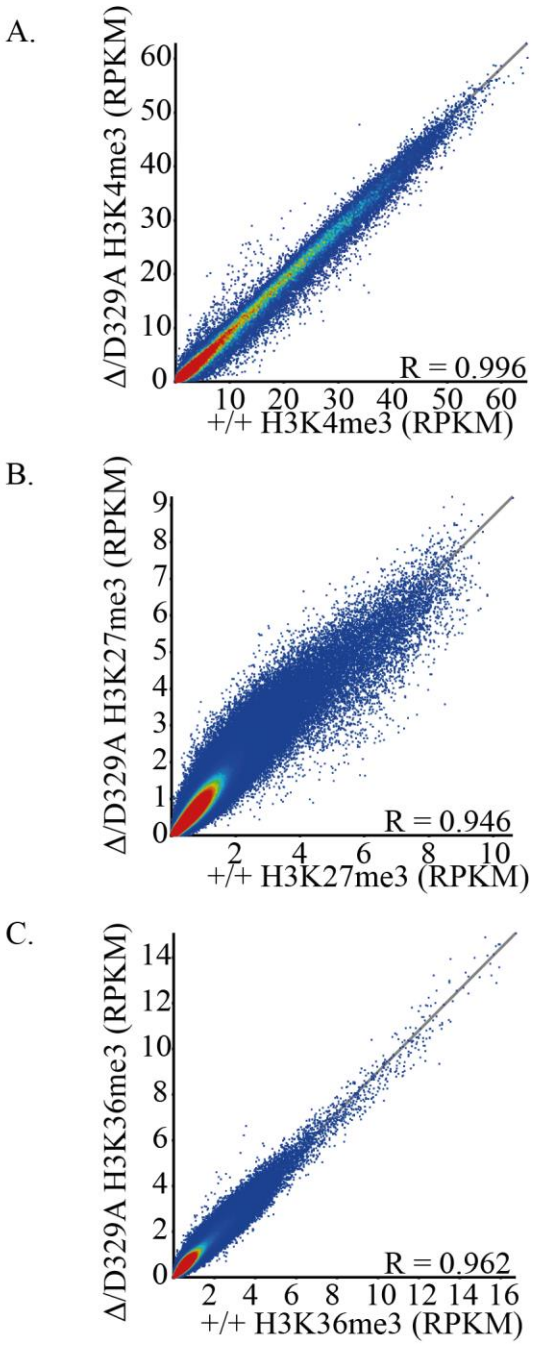


Figure 6.7. Distribution of histone PTM enrichment in hypothalamus. Scatterplots showing correlation between enrichments for the following histone modifications between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ hypothalamus: **A.** H3K4me3; **B.** H3K27me3; **C.** H3K36me3. Tiles of 2kb with a step of 1kb were used for analyses.

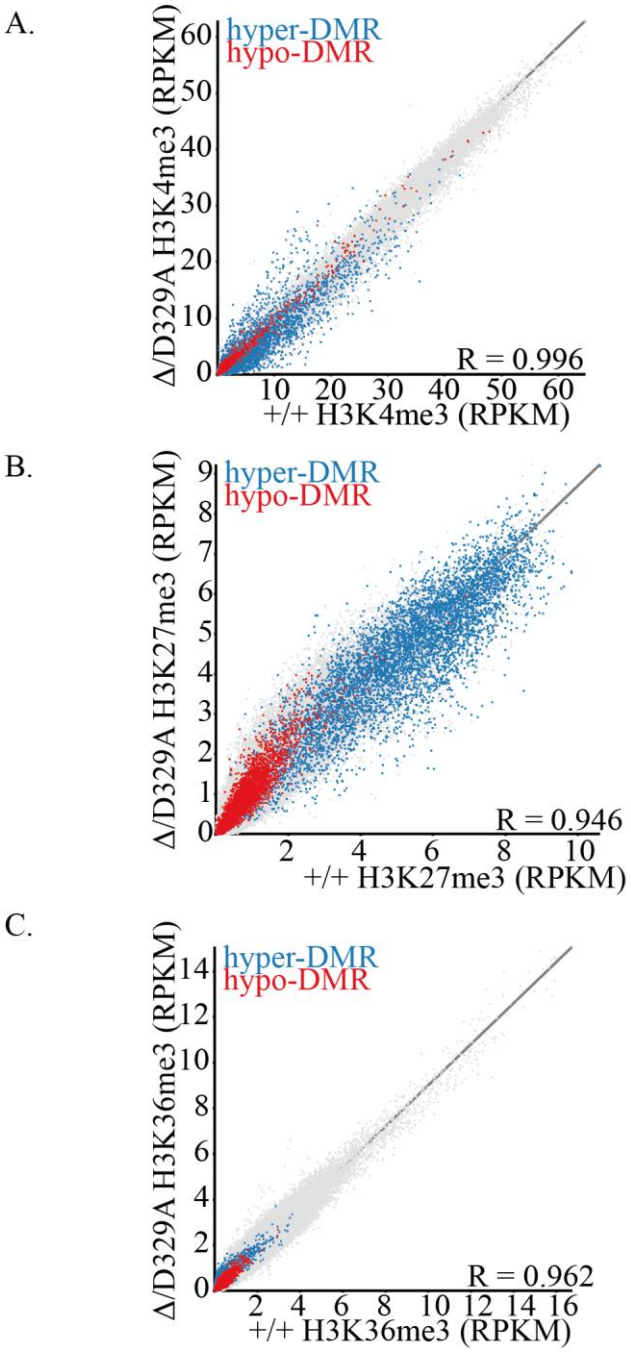


Figure 6.8. Link between hypothalamus DMRs and histone PTMs. Scatterplots showing correlation between enrichments for **A.** H3K4me3; **B.** H3K27me3; **C.** H3K36me3 histone modifications between *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$ hypothalamus with those overlapping hypothalamus DNA methylation DMRs colour based on their status. Tiles of 2kb with a step of 1kb were used for analyses.

MACS peak caller was used to call peaks in H3K4me3, H3K27me3 and H3K36me3 ChIP-seq datasets. Peaks are areas showing a clear histone modification enrichment in

comparison to surrounding chromatin. Hypothalamus DMRs were labelled based on which histone PTM peak they overlapped. The same random tile list was used as previously to represent general genome trends. I found that 70% of hyper-DMRs were at bivalent chromatin and an additional 26% at H3K27me3 (Figure 6.9), therefore 96% of all hypermethylated domains were marked by H3K27me3. Only 18.3% of hypomethylated and 19.45% of random tiles fell over bivalent and H3K27me3 marked chromatin. In contrast, only 1.8-2.1% of hypo- or hyper-DMRs overlapped the H3K36me3 mark, again suggesting that the activity of DNMT3A^{D329A} is not linked to H3K36me3.

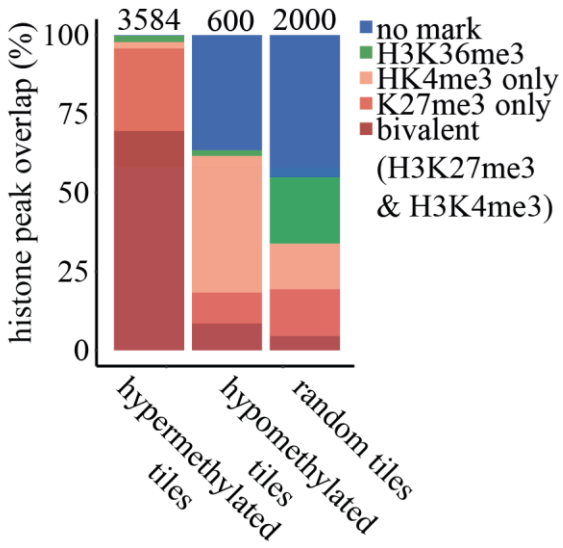


Figure 6.9. Hypothalamus DMR and histone PTM overlap. Percentages of hypermethylated and hypomethylated DMRs (100-CpG) which overlap bivalent (H3K4me3 and H3K27me3), H3K27me3 only, H3K4me3 only, H3K36me3 peaks or none. A random set of 2000 tiles (excluding DMR tiles) was used as a whole genome representative. Tiles that show overlap with H3K36me3 and H3K4me3 or H3K27me3 were included in H3K36me3 group. Peaks were called using MACS peak caller in adult hypothalamus ChIP-seq.

Aberrant DNA methylation targeting to methylation canyons marked by H3K27me3 could be explained by two hypotheses. The first hypothesis would be that the D329A mutation ablates the recruitment of DNMT3A to H3K36me3, therefore leading to

redistribution of DNMT3A^{D329A} to other random genomic regions. Subsequently, mutant DNMT3A could methylate those random regions. This aberrant methylation would most likely be detected as significant in typically unmethylated places. An alternative hypothesis would be that H3K27me3 is actively recruiting DNMT3A^{D329A}, which leads to hypermethylation of the genome compartments enriched by this histone mark.

To test the first hypothesis and identify whether there was loss of DNA methylation at regions enriched for H3K36me3, I analysed gene bodies. Gene bodies were separated into groups of those showing low- (bottom quintile) or high- (top quintile) levels of H3K36me3. There were no differences, apart those observed globally, in levels of DNA methylation over gene bodies in any of the genotypes (Figure 6.10).

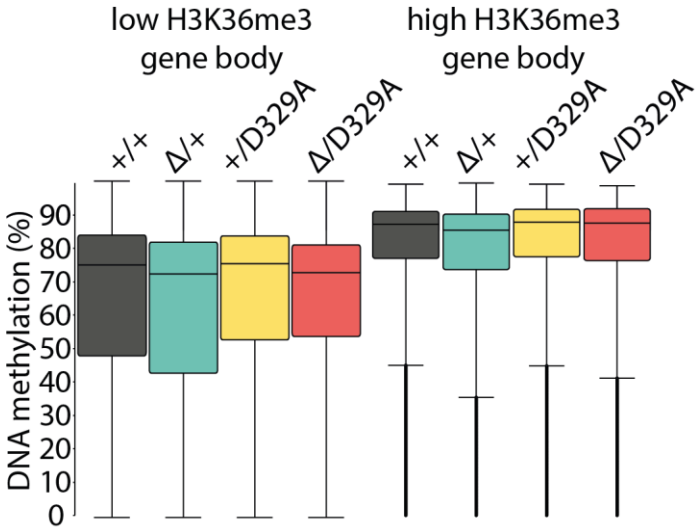


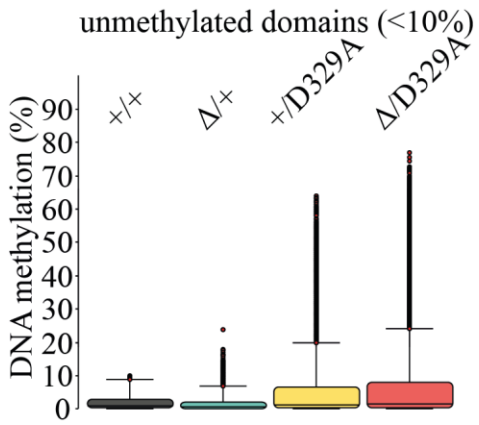
Figure 6.10. DNA methylation levels over H3K36me3 marked gene bodies. Gene bodies were separated based on levels of H3K36me3. Low – bottom quintile, high – top quintile. Kruskall-Wallis rank sum test results in p-value=4.1e-15 and is significant for all post-hoc pairwise Wilcoxon rank sum tests, with exception of *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$ pair, where p=0.336.

In order to see whether aberrant methylation gain was targeted to all of the normally unmethylated regions, or those marked by H3K27me3 specifically, I selected tiles showing <10% DNA methylation in *Dnmt3a*$^{+/+}$ adult hypothalamus (4.6% of the genome). Unmethylated domains were then separated into seven categories: 1) low-

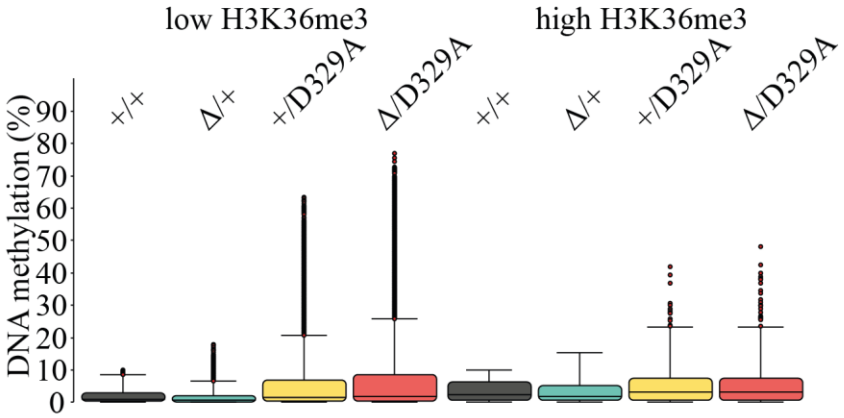
H3K4me3 (excl. bivalent), 2) high-H3K4me3 (excl. bivalent), 3) low-H3K36me3, 4) high-H3K36me3, 5) low-H3K27me3 (excl. bivalent), 6) high-H3K27me3 (excl. bivalent), 7) bivalent (high-H3K4me3 and high-H3K27me3), where low- and high-enrichment is top and bottom quintile, respectively.

The levels of DNA methylation across all four genotypes were very similar in all unmethylated domains (Figure 6.11a) with low- or high- H3K36me3 and H3K4me3 levels (Figure 6.11b and Figure 6.11c). DNA methylation was unaffected in regions with low-H3K27me3 too (Figure 6.11d). However, regions that were highly enriched in H3K27me3 showed much higher levels of DNA methylation in *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$ hypothalamus. This increase in DNA methylation was also observed in regions marked by bivalent chromatin, although to a lesser extent. Thus, hypomethylation of genomic domains *per se* is not sufficient to recruit and activate DNMT3A^{D329A}. Meanwhile, a lower gain in DNA methylation at bivalent regions suggests that the presence of H3K4me3 remains antagonistic to DNMT3A^{D329A} activity. Together, these results support the conclusion that DNMT3A^{D329A} preferentially relocalises to H3K27me3-marked chromatin.

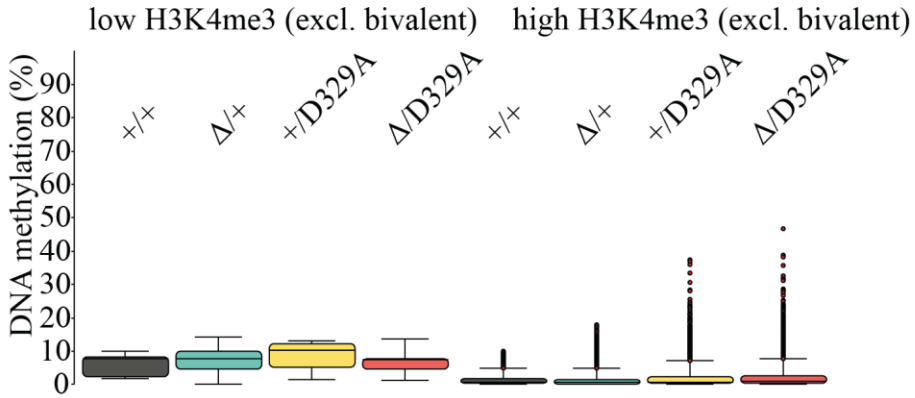
A.


unmethylated domains (<10%)
DNA methylation (%)
+/+
Δ/+
+/D329A
Δ/D329A

B.


low H3K36me3
high H3K36me3
DNA methylation (%)
+/+
Δ/+
+/D329A
Δ/D329A

C.


low H3K4me3 (excl. bivalent)
high H3K4me3 (excl. bivalent)
DNA methylation (%)
+/+
Δ/+
+/D329A
Δ/D329A

D.

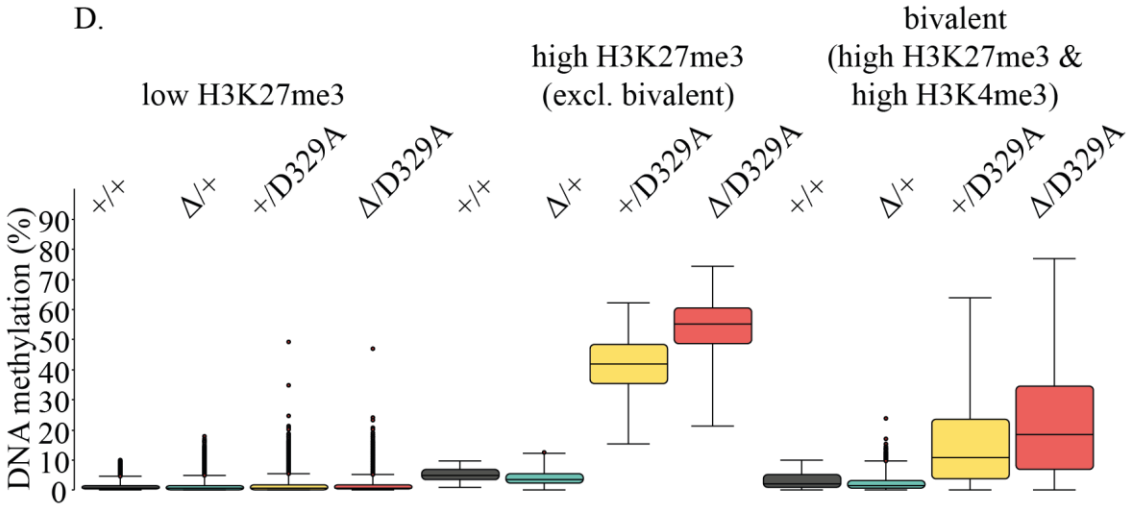

low H3K27me3
high H3K27me3
(excl. bivalent)
bivalent
(high H3K27me3 &
high H3K4me3)
DNA methylation (%)
+/+
Δ/+
+/D329A
Δ/D329A

Figure 6.11. DNA methylation changes at the normally unmethylated genomic fraction stratified by histone PTM status. DNA methylation levels across **A.** unmethylated genomic regions (<10% DNA methylation in wild-type hypothalamus) amongst different genotypes, over unmethylated regions overlapping low and high **B.** H3K36me3, **C.** H3K4me3, **D.** H3K27me3 and bivalent chromatin. Low level - bottom quintile (0-20%) of enrichment, high level – top quintile (80-100%) of enrichment. Histone mark enrichment was quantitated over 100-CpG DMR tiles. In order of appearance: Kruskall-Wallis rank sum test results were **A.** p-value <2.2e-16 and post-hoc pairwise Wilcoxon rank sum tests were significant. **B.** low K36me3, Kruskall-Wallis p-value <2.2e-16 and pairwise all significant. High H3K36me3 Kruskal-Wallis was p=5.583e-10, significant for all pairs, except *Dnmt3a*$^{+/D329A}$ and *Dnmt3a*$^{\Delta/D329A}$. **C.** low H3K4me3 Kruskall-Wallis p=0.571, no significant pairs. High H3K4me3, Kruskall-Wallis p<2.2e-16, and all pairwise significant. **D.** low H3K27me3 Kruskall-Wallis p<2.2e-16, pairwise comparisons not significant for pairs *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{+/D329A}$, *Dnmt3a*$^{+/+}$ and *Dnmt3a*$^{\Delta/D329A}$. High H3K27me3 and bivalent, Kruskal-Wallis p<2.2e-16 and all pairwise comparisons significant.

6.2 Altered chromatin environment at DMR sites

Because H3K27me3 and DNA methylation show mutual exclusivity in most of the genomic contexts (Lindroth et al., 2008; Bartke et al., 2010; Brinkman et al., 2012), I was interested in whether the histone landscape would be altered, with DNA methylation replacing H3K27me3 at hypermethylated DMRs. To do this I calculated the net change in DNA methylation levels (%) between *Dnmt3a*$^{\Delta/D329A}$ and *Dnmt3a*$^{+/+}$ and compared it to the change in enrichment (RPKM) of H3K27me3 over all DMR regions. The levels of DNA methylation were marginally associated with change in H3K27me3. Areas showing the higher net methylation gain had a tendency to show lower H3K27me3 enrichment in *Dnmt3a*$^{\Delta/D329A}$ samples (Figure 6.12a), but there was no major loss of this histone mark, as could have been predicted. Therefore, these marks appeared to co-exist with hypermethylated domains retaining a substantial amount of H3K27me3 (Figure 6.13). Interestingly, I observed that the extent of DNA methylation gain correlated much better with initial enrichment of H3K27me3 over the locus in *Dnmt3a*$^{+/+}$, than its loss, and therefore high levels of H3K27me3 would be predictive of a higher DNA methylation gain (Figure 6.12b).

Finally, I used the ChIP-seq data to assess chromatin changes at DEGs. I found that for a subset of DEGs the increase in expression was positively linked with changes in promoter H3K4me3 (Figure 6.12c). Together, we see that presence of the DNMT3A^{D329A} protein results in gain of DNA methylation directly, but indirectly also causes a redistribution of chromatin marks at those sites. When assessed all together, I can observe that gain of DNA methylation sometimes led to increased gene expression, increase in promoter H3K4me3, decrease in H3K27me3 and increase of H3K36me3 over the upregulated gene body (Figure 6.13). These findings support the link between H3K27me3 and abnormal localisation of DNMT3A^{D329A}, and suggest that the normal histone landscape is affected by the aberrant gain in DNA methylation.

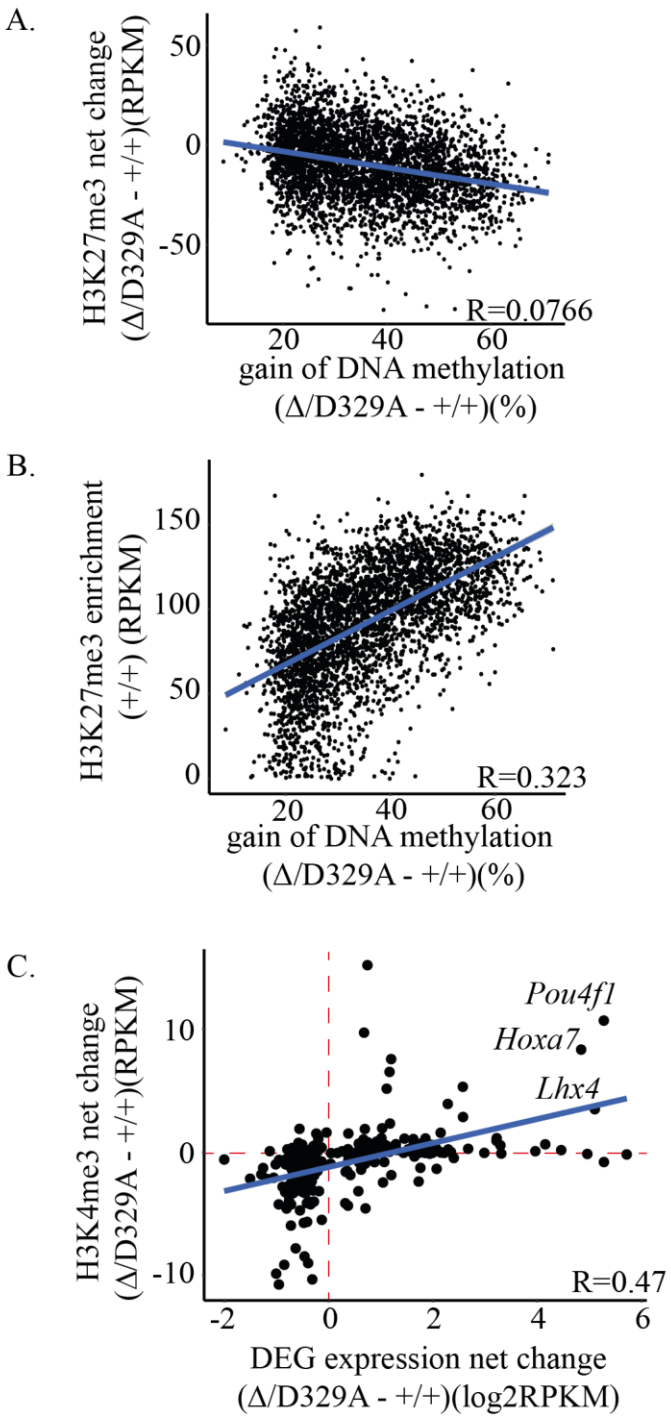


Figure 6.12. Histone PTM and DNA methylation gain associations. Scatterplots showing the relationship between **A.** gain of DNA methylation in *Dnmt3a*$^{\Delta/D329A}$ hypothalamus and change in enrichment of H3K27me3, **B.** relationship between levels of H3K27me3 enrichment at *Dnmt3a*$^{+/+}$ and DNA methylation gain, and **C.** change in DEG expression and change in respective gene promoter H3K4me3 enrichment between *Dnmt3a*$^{+/+}$ and *Dnmt3a* $^{\Delta/D329A}$.

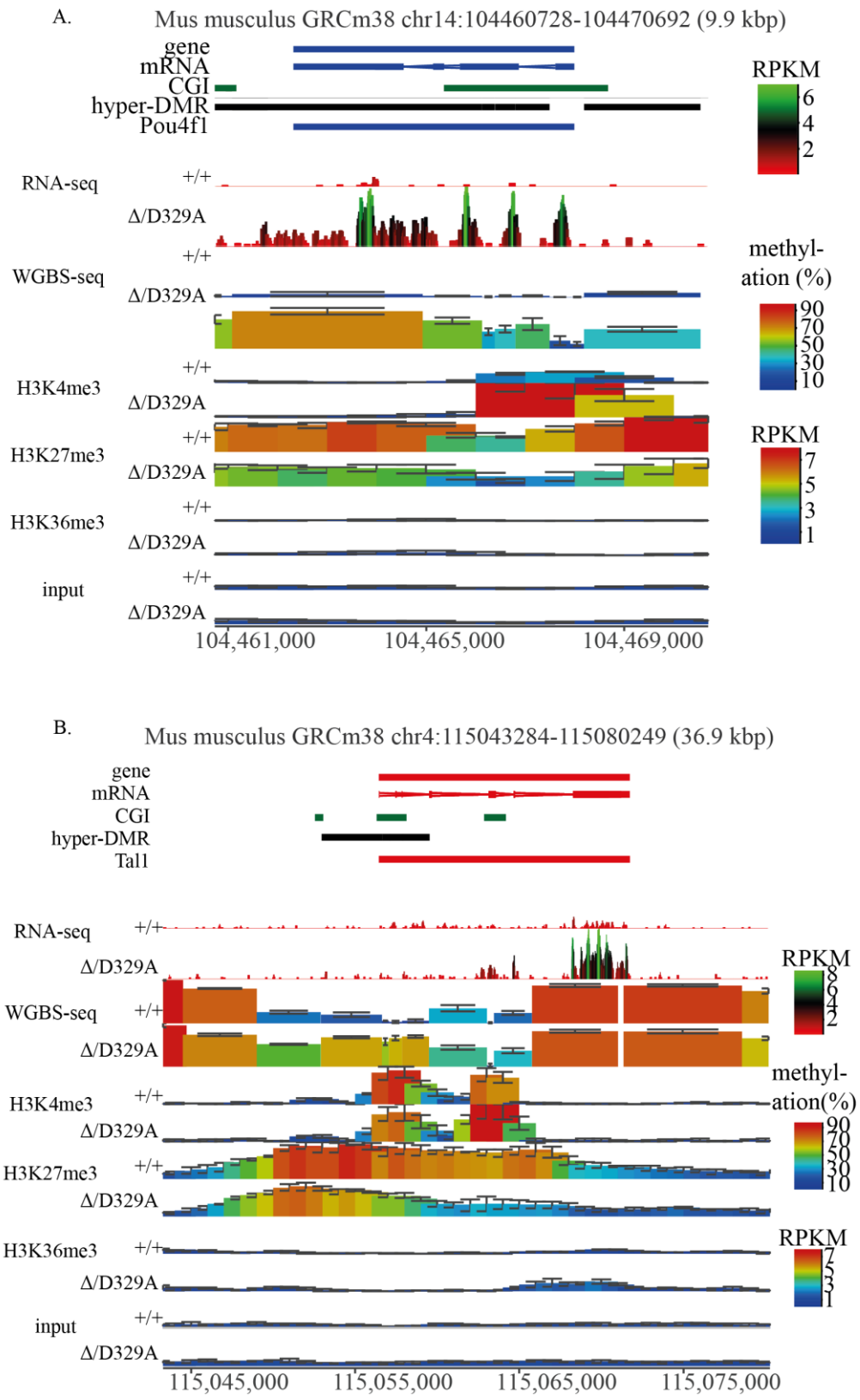


Figure 6.13. D329A mutation caused genomic landscape changes. Representative genome browser views of gene expression, DNA methylation, and distribution of H3K4me3, H3K27me3 and H3K36me3 marks in adult male hypothalamus. H3K4me3 and H3K36me3 are enriched over actively transcribed CGI promoters and gene bodies,

respectively. H3K27me3 shows a broad enrichment over transcriptionally silent genes and a decrease of enrichment upon methylation gain. Genes **A.** *Pou4f1* and **B.** *Tal1* are up-regulated and hypermethylated in *Dnmt3a*$^{A/D329A}$. Each colour-coded block represents a 50bp window with 10bp step size in the gene expression dataset; a 100-CpG window in the methylation dataset; and 2kb window with 1kb step size in the histone enrichment datasets. For gene and mRNA tracks, the colour indicates direction, where red is a forward strand and blue is a reverse strand. Error bars indicate standard deviation.

6.3 Summary and discussion

Together, findings in this chapter suggest that the function of DNMT3A *in vivo* is not necessarily reliant on its association with H3K36me3. Instead, I see that DNMT3A^{D329A} targets H3K27me3-marked chromatin specifically. This gain in DNA methylation is strongly associated with initial levels of H3K27me3 but also subsequently led to some loss of H3K27me3 over hypermethylated domains. Similarly, genes showing increased expression upon hypermethylation have increased levels of H3K4me3 at their promoters.

The original study showed that the DNMT3A-PWWP domain interacted with H3K36me3 and the D329A mutation ablated this binding (Dhayalan et al., 2010). Therefore, it is rather surprising that I did not find any evidence of DNA methylation changes linked to H3K36me3 enrichment. Again, it highlights the discrepancies of studies conducted *in vitro* and *in vivo*, and raises the possibility that the PWWP domain is not interacting with this histone mark or, alternatively, the D329A mutation does not have the same ablation effect *in vivo.*

By systematically evaluating the histone PTM context, I found that DNA methylation is not leaking into all unmethylated fractions of the genome, but rather is specifically targeted to the regions enriched in H3K27me3. Therefore, another possible explanation would be that it results in altered binding conformation that allows interaction with H3K27me3, observed as a target of DNMT3A^{D329A}. Generally, DNA methylation and H3K27me3 are considered to be mutually exclusive, at least at DNA methylation valleys (Bogdanović et al., 2011; Brinkman et al., 2012; Murphy et al., 2013; Statham et al., 2012). Thus far, there is no evidence of a direct interaction, binding or repression, between DNMT3A or H3K27me3 (Dhayalan et al., 2010; Manzo et al., 2017; Mauser et

al., 2017). Nonetheless, DNMT3A has been found to associate with shores of DNA methylation valleys, so it might already be located closed to bivalent chromatin and have a weak affinity to H3K27me3 (Gu et al., 2018; Manzo et al., 2017). Given that H3K27me3 is established by PRC2, and DNMT3A is found as a direct interactor of PRC2 proteins (Manzo et al., 2017), it is possible that a conformation change caused by the D329A mutation somehow enhances this interaction or leads to further recruitment of mutant DNMT3A to H3K27me3 domains. Similarly, it has been reported that PRC2 member EZH2 can directly recruit DNMTs to the genome (Viré et al., 2006).

Although H3K27me3 and DNA methylation are not generally found together, DNA methylation gain at domains marked by H3K27me3 is found in cancer, where the balance is possibly tipped, and altered cellular environment allows it (Schlesinger et al., 2007; Gao et al., 2014; Takeshima et al., 2015). Similarly, DNA methylation and H3K27me3 work in concert to silence X chromosome (Galupa and Heard, 2015). The AEBP2 protein of PRC2 was shown to target PRC2 to CG-rich and methylated DNA (Gao et al., 2014; Wang et al., 2017). Therefore, PRC2 could maintain the H3K27me3 even in the presence of aberrant DNA methylation gain. Conversely, a different study showed that PRC2 is involved in an insular chromatin formation at DNA methylation canyons and therefore promotes hypomethylation of these domains (Li et al., 2018). The tight interplay between DNA methylation and the H2K27me3 mark is potentially modulated by other chromatin interacting factors. Together they confer environments where there is mutual exclusivity and where permissiveness to co-exist is determined.

A caveat of whole tissue analysis, especially when relatively low changes in histone modification enrichment are observed, is heterogeneity of the tissue. If changes in DNA methylation do not occur in all cells, the limited change in H3K27me3 levels could simply be a mixture of signals coming from cells that are unaffected and those with a strong change. A more definitive way to answer this question would be to conduct a sequential ChIP-bisulphite sequencing, where the genome fraction marked by histone modification is first pulled down and then DNA methylation levels over those domains are assessed.

This chapter shows that aberrant gain of DNA methylation is targeted specifically to the chromatin marked by H3K27me3 and, to a lesser extent, bivalent chromatin, containing both H2K27me3 and H3K4me3.

7 Generation of a bioDNMTA-D329A ES cell line for mechanistic studies

Declaration of author contributions

The majority of the work described in this chapter was conducted by the author while on an academic visit to Tuncay Baubec's laboratory (Department of Molecular Mechanisms of Disease, University of Zurich). Joel Wirz kindly helped to conduct the Co-IP-MS experiment and ran the MS machine.

The limitations of investigating the DNMTA^{D329A} protein *in vivo* include the small quantity of adult hypothalamus tissue, which prohibits many biochemical approaches, the cellular complexity of tissue samples, which can confound interpretations, the complex whole-body phenotype, and lack of biological tools, such as antibodies suitable for immunoprecipitation. These constraints made it challenging to study the impact of the D329A mutation on the biochemical properties of the DNMT3A protein. A group led by Dirk Schübeler established an elegant recombinase-assisted mapping of biotin-tagged proteins (RAMBiO) system in which DNA binding proteins can be tagged individually without disrupting protein structure (Baubec et al., 2013). Tagged genetic constructs can be inserted into a specific locus in the mouse ESC genome with a known genetic background, and robust antibodies or resins can be used to obtain DNMT binding maps (ChIP-seq) and identify co-factors/interactors (co-Immunoprecipitation).

A schematic representation of the generation of ESCs containing a biotinylated protein of interest, bioDNMT3A1^{D329A}, is shown below (Figure 7.1). Briefly, ESCs that contain a genetically engineered site for recombinase-mediated cassette exchange (RMCE) and constitutively express biotin ligase *BirA* are used. The target site lies at the *β-globin* locus and contains two fused selection elements, hygromycin-resistance (*Hy*) and Herpes Simplex Virus-1 thymidine kinase ganciclovir-sensitivity (*HSV-Tk*) genes, flanked by inverted *loxP* elements (Lienert et al., 2011). Constructs that contain an *AviTag* element

and gene of interest are introduced to cells together with *Cre* recombinase. Cre recombinase would drive RMCE and the original selection cassette would be exchanged by a transgene carrying cassette. Once the transgene is expressed, the *AviTag* sequence at the N-terminus of the gene of interest is translated into a short peptide that is a substrate of an *E. coli* biotin ligase BirA. BirA covalently conjugates free-floating biotin molecules to the AviTag peptide making the transgenic protein biotinylated (Beckett et al., 1999). Binding of biotin and (strept)avidin is one of the strongest non-covalent molecular interactions and many molecular products, such as antibodies and resins, are created to harness this unique property. Here, this system was used to create *bioDnmt3a1*^{D329A} cell line for further work.

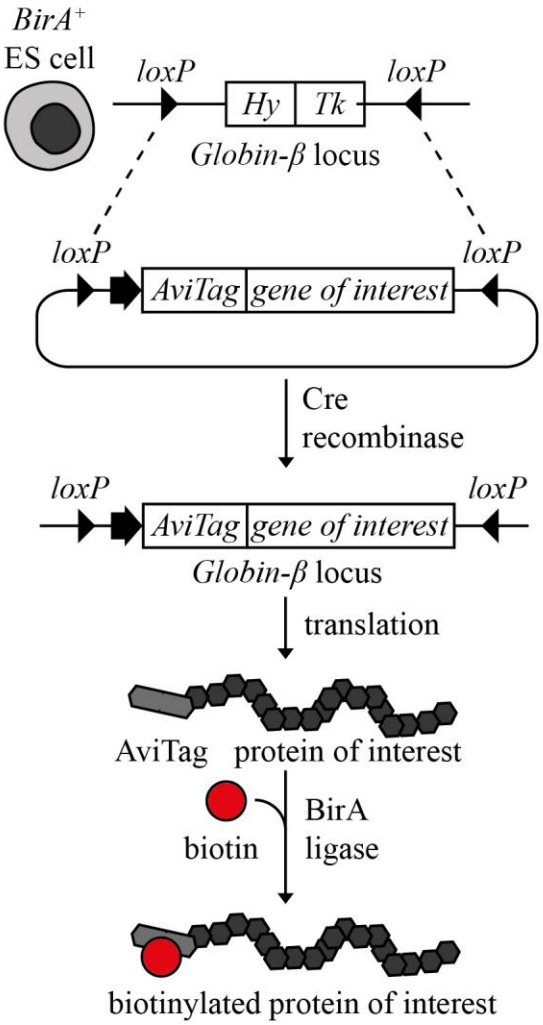


Figure 7.1. Schematic representation of generation of ES cells containing biotinylated protein of interest. Stem cells which constitutively express biotin ligase *BirA* and contain a selection cassette with hygromycin-resistance (*Hy*) and Herpes Simplex Virus-1 ganciclovir-sensitivity thymidine kinase (*Tk*) genes, flanked by inverted *loxP* elements (indicated by triangles), at *Globin-β* locus are used. A plasmid containing a promoter (indicated by black arrow), BirA target *AviTag* sequence, and gene of interest flanked by inverted *loxP* elements are introduced together with *Cre* recombinase. Cre drives recombinase-mediated cassette exchange (RMCE) and the desired construct is inserted into the endogenous locus. Once the gene of interest is translated, BirA would covalently conjugate free-floating biotin molecule to the AviTag peptide. Figure modified from Baubec et al. (2013).

7.1 Generation of *bioDnmt3a*^{D329A} ES cell lines

First, the D329A mutation was introduced into the pJET-D3A1 plasmid, which contains *Dnmt3a* cDNA, by site-directed mutagenesis. *Dnmt3a1*^{D329A} was cut out of pJET-D3A1-D329A plasmid and ligated into mammalian expression vector pCAGbioTEVeGFP in place of *eGFP* cDNA (both plasmids provided by T. Baubec, UZH Switzerland). Correct insertion and absence of spontaneous mutations were assessed by Sanger sequencing. The resulting pCAGbioTEV-DNMT3A1_D329A, together with *Cre* recombinase expression plasmid pIC-Cre (Gu et al., 1993), were introduced into HA36CB1 recipient ES cells, henceforth referred to as *wild-type*, by electroporation (Baubec et al., 2013). HA36CB1 cells already constitutively express BirA ligase and have the RMCE selection locus. After electroporation, cells were subjected to ganciclovir selection. Individual clones were picked and genotyped to confirm insertion of *bioDnmt3a1*^{D329}. Four individual clones were assessed by western blot and all four showed presence of biotinylated DNMT3A^{D329A} protein (Figure 7.2).

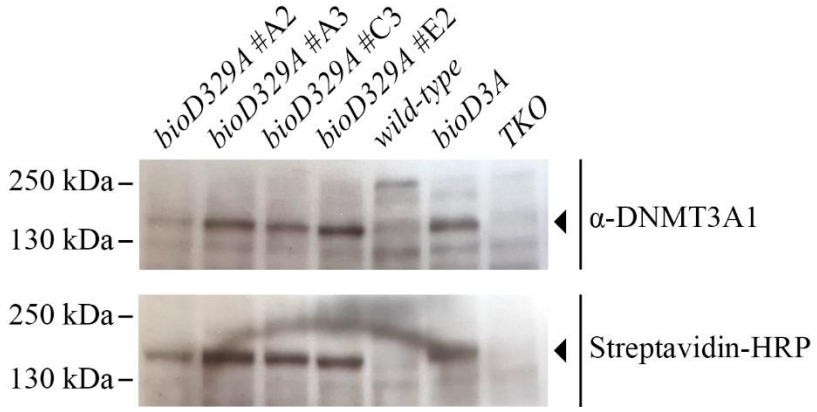


Figure 7.2. Western blot of transgenic ES cell nuclear extracts for biotinylated protein detection. First, primary antibody against DNMT3A1 was used for endogenous protein detection. Blot was then stripped and HRP-conjugated streptavidin was used to assess biotinylated protein presence. Individual clones represent *bioDnmt3a1*^{D329A} HA36CB1 cell lines; wild-type – HA36CB1 cell line, bioD3A – *bioDnmt3a1* HA36CB1 cell line (positive control) and TKO – DNMT TKO-133 *Dnmt* triple knock-out cell line (negative control) (the latter two provided by T. Baubec, UZH Switzerland). Arrows point to the position of bands marked by the antibody or streptavidin-HRP. 20µg of nuclear lysate was loaded per well.

7.2 Assessment of bioDNMT3A^{D329A} binding factors using Co-IP-MS

In order to identify factors that bind bioDNMT3A and/or bioDNMT3A^{D329A}, biotin co-immunoprecipitation followed by mass spectrometry was done. First, I performed extraction of nuclear fractions, followed by co-immunoprecipitation using streptavidin coated beads. Pulled-down fractions were run on a mass-spectrometer and identified peptides were analysed. Joel Wirz kindly helped me with processing the samples and ran them on a mass-spectrometer. The nuclear fraction of ESCs was extracted using hypotonic buffers, biotinylated complexes were immunoprecipitated using streptavidin beads. Protein cleavage to peptides was done with trypsin, using FASP digestion method (Wiśniewski et al., 2009). Protein identification from MS-raw data was performed and mapped to the mouse proteome using MaxQuant label-free quantification. The analysis was done using *Proteus* package in R statistical software (Gierlinski et al., 2018).

The number of peptides found in each replicate suggested a successful immunoprecipitation, since the wild-type cell line had a lower number of peptides found than either *bioDnmt3a1* or *bioDnmt3a1*^{D329A} lines, despite equivalent input amounts (Figure 7.3a). Pairwise Pearson's correlation analysis showed high similarity between *bioDnmt3a1* and *bioDnmt3a1*^{D329A} clones, with exception of sample *bioDnmt3a1#2* (Figure 7.3b). It showed lower peptide count than other samples of the same genotype and separated in PCA (Figure 7.3c). Given these findings, sample *bioDnmt3a1#2* was excluded from further analysis. The adjusted PCA plot indicated a clear separation of *wild-type* samples, meanwhile *bioDnmt3a1* and *bioDnmt3a1*^{D329A} clones clustered together (Figure 7.3d).

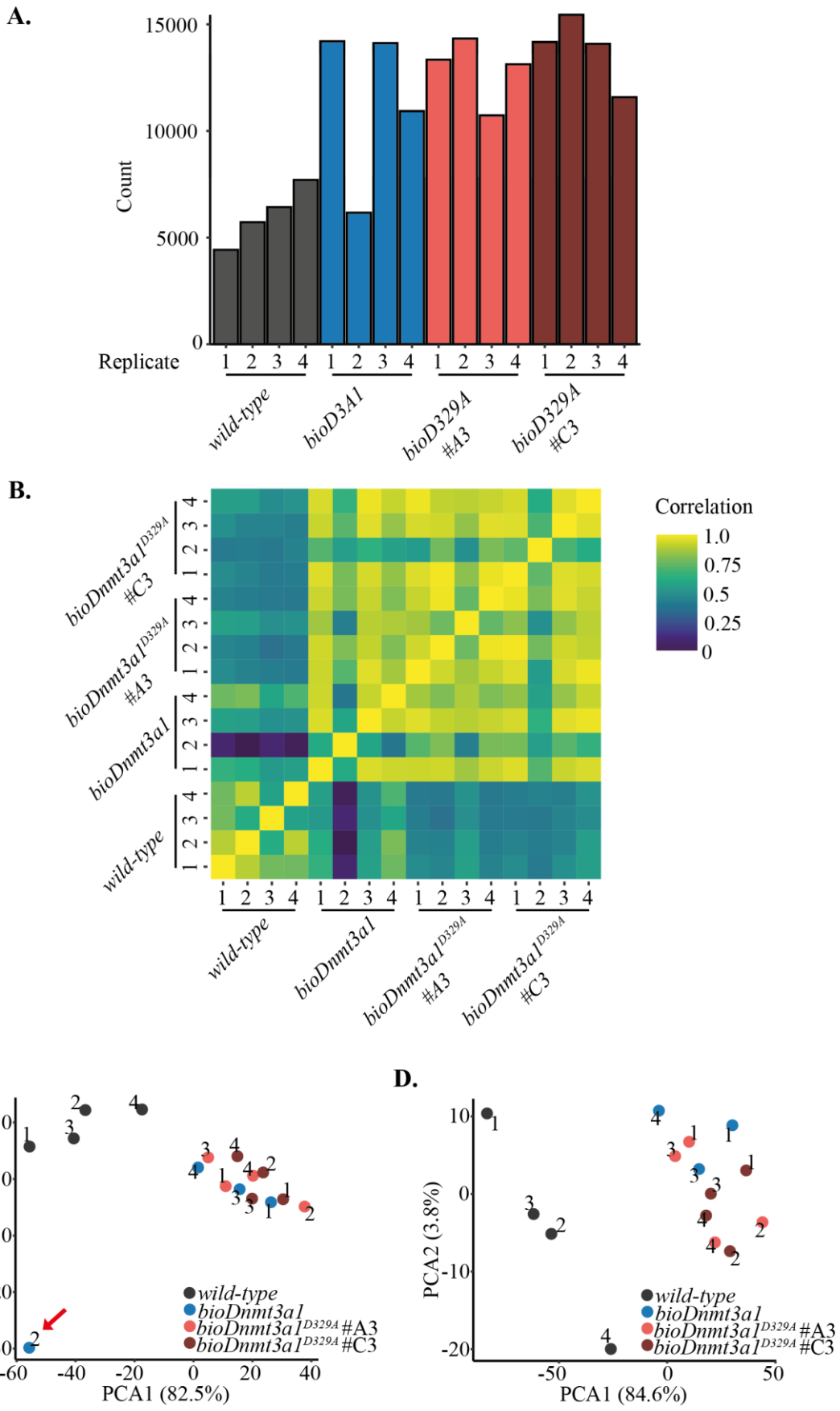


Figure 7.3. Co-IP-MS sample quality assessment. A. Total count of peptides identified by MaxQuant in each replicate of biotin Co-IP-MS. **B.** Pearson's pairwise correlation

levels between all Co-IP-MS samples. Numbers represent replicates. **C.** PCA of all samples. Notably, replicate 2 of *bioDnmt3a1* cell line shows comparatively lower correlation with all other samples in panel A and is separated by PCA in panel B (indicated by the red arrow). **D.** PCA of samples with exclusion of replicate 2 of *bioDnmt3a1* cell line.

Peptides were then aggregated and assigned to proteins using the Leading Razor Protein method (Silva et al., 2006). Data were normalised to match median intensities across all samples (Figure 7.4). The statistical tool limmaDE was used to find differential representation of proteins in a pairwise manner, focusing on *bioDnmt3a1* against *bioDnmt3a*^{D329A} clones #A3 and #C3. Plots of fold-change and *p*-value for these comparisons can be found below (Figure 7.5). However, after Benjamini-Hochberg multiple comparison adjustments, there were no proteins significantly overrepresented in *bioDnmt3a*^{D329A} clone #A3 and/or #C3 pull-downs when compared to *bioDnmt3a1*.

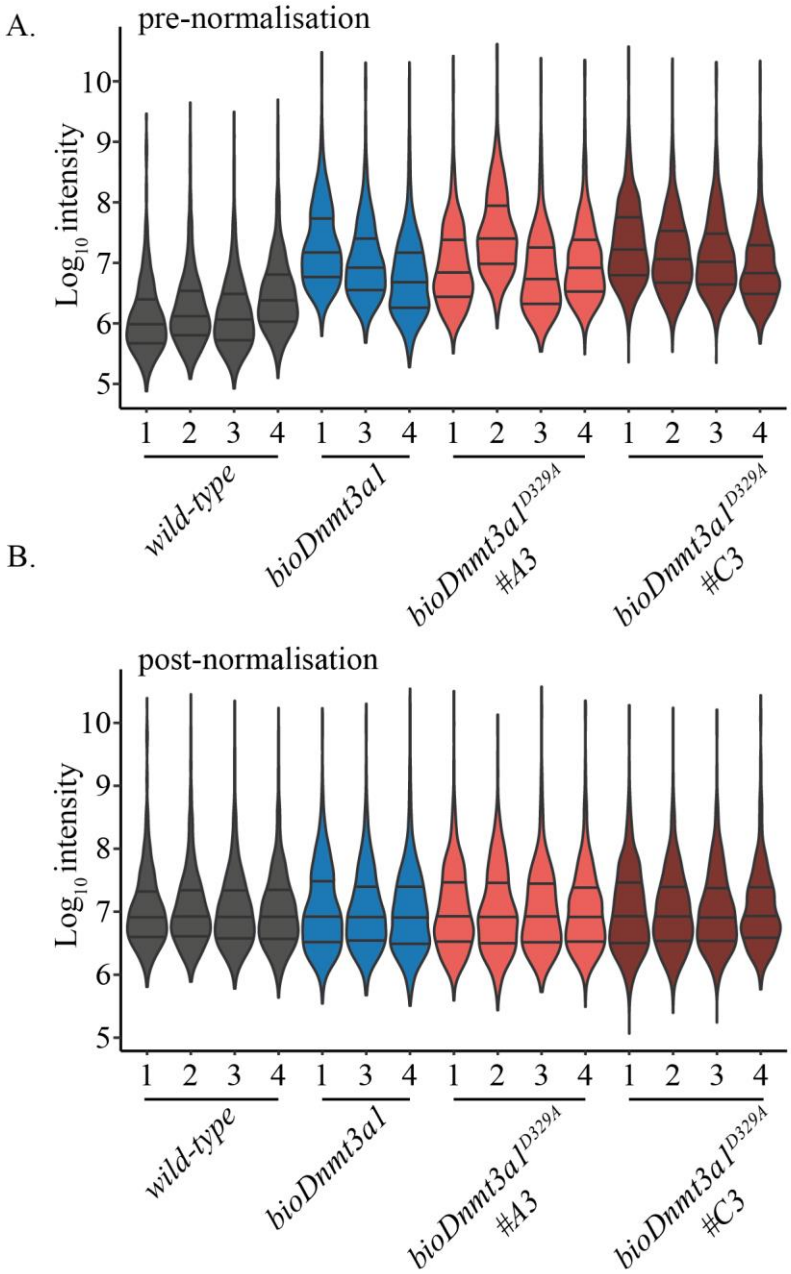


Figure 7.4. Co-IP-MS sample intensity value normalisation by mean. Panel **A.** shows $\log_{10}$ values of peptide intensities before normalisation and panel **B.** shows normalised intensity values. Numbers refer to replicates.

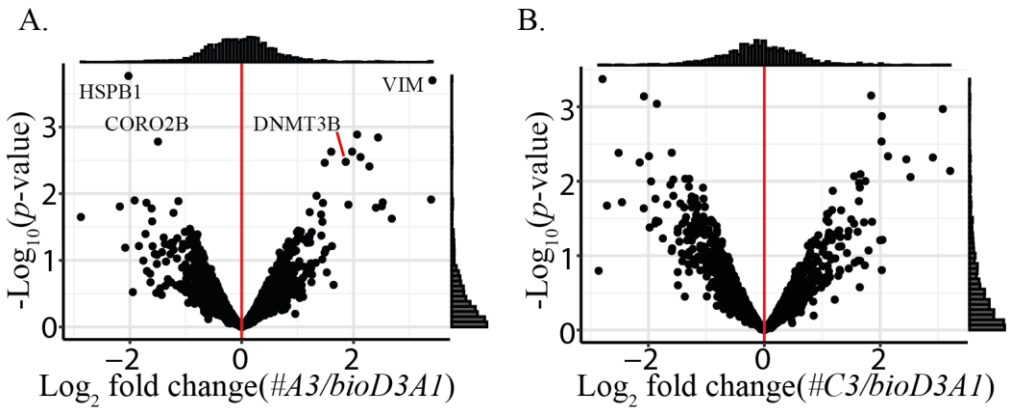


Figure 7.5. Volcano plots showing differential representation of proteins in *bioDnmt3a1*^{D329A} clones #A3 and #C3 in comparison to *bioDnmt3a1*. Histograms indicate distributions of either $\log_2$ fold change of proteins or $-\log_{10}$ *p*-values.

There were 197 differentially enriched proteins in *bioDnmt3a1* compared to *wild-type*, providing a list of potential DNMT3A1 interactors in ESCs. The top 20 hits showing enrichment in *bioDnmt3a1* samples in comparison to *wild-type* are listed in Appendix 3. It was reassuring to find DNMT3A ($\log_2$FC = 3.6) and DNMT3L ($\log_2$FC = 3.7) in this list, both of which were found to be significantly enriched in *bioDnmt3a1* and mutant cells lines (Figure 7.6 and Figure 7.7). This assured that the IP indeed pulled-down known DNMT3A-binding factors, and suggested that the N-terminal end should be unaffected by the biotin tag. Methyl-binding domain 3 (MBD3) protein ($\log_2$FC = 3.8), known to bind DNMT3A (Datta et al., 2005), was also in the top 20 list (Figure 7.8). DNMT3B ($\log_2$FC = 2.6) ranked 57 in the list sorted based on significant fold-change.

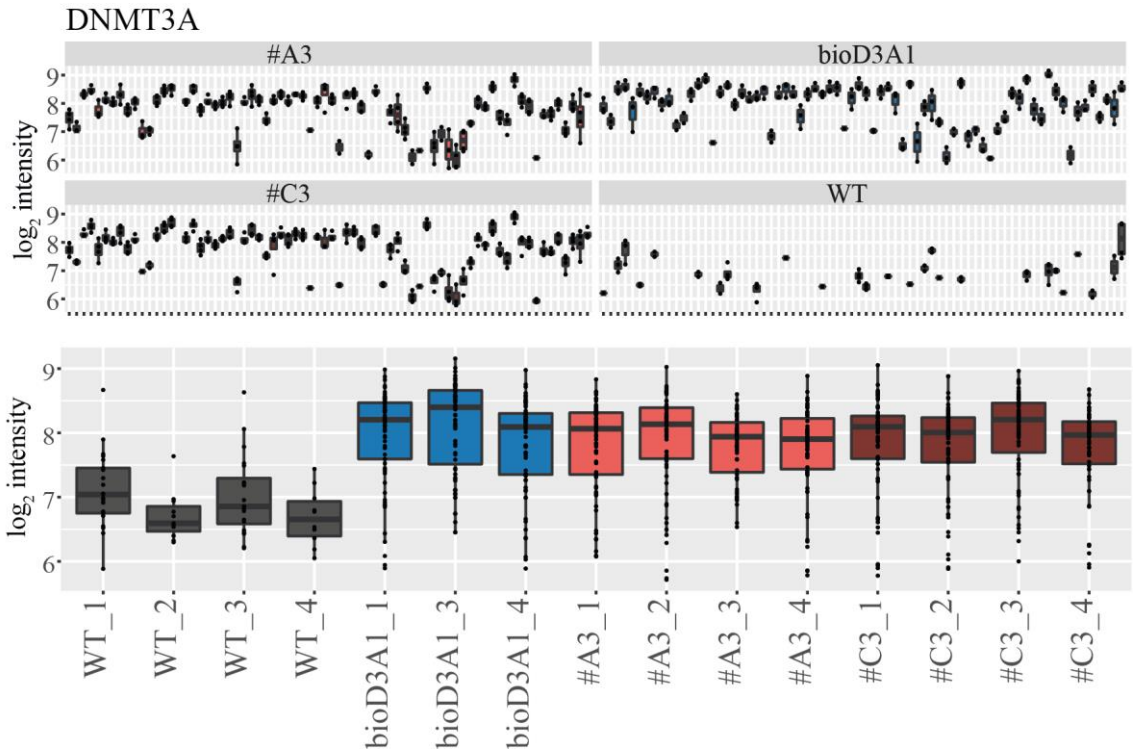


Figure 7.6. Peptide data for DNMT3A. Top panel shows individual peptide intensities per genotype group, where every box represents a unique peptide sequence along the protein; peptides are displayed in alphabetical order based on amino acid sequence. Bottom panel shows peptide intensity per individual sample. A single dot represents one detected peptide.

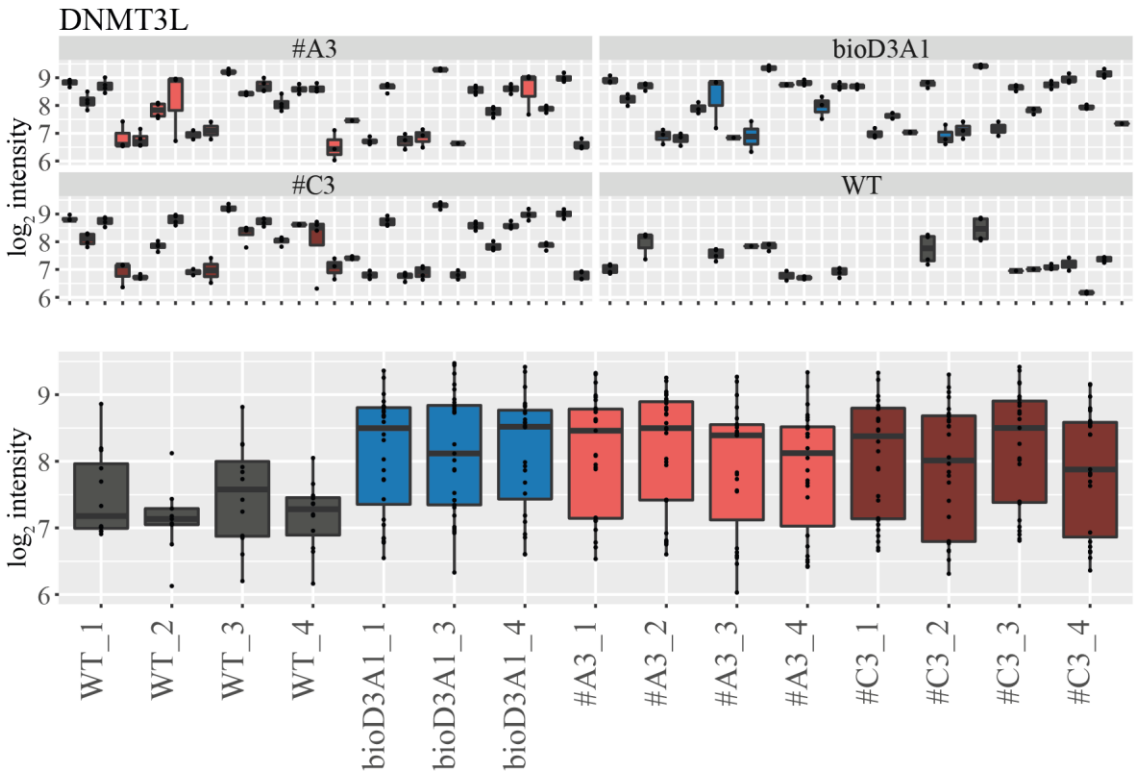


Figure 7.7. Peptide data for DNMT3L. Top panel shows individual peptide intensities per genotype group, where every box represents a unique peptide sequence along the protein; peptides are displayed in alphabetical order based on amino acid sequence. Bottom panel shows peptide intensity per individual sample. A single dot represents one detected peptide.

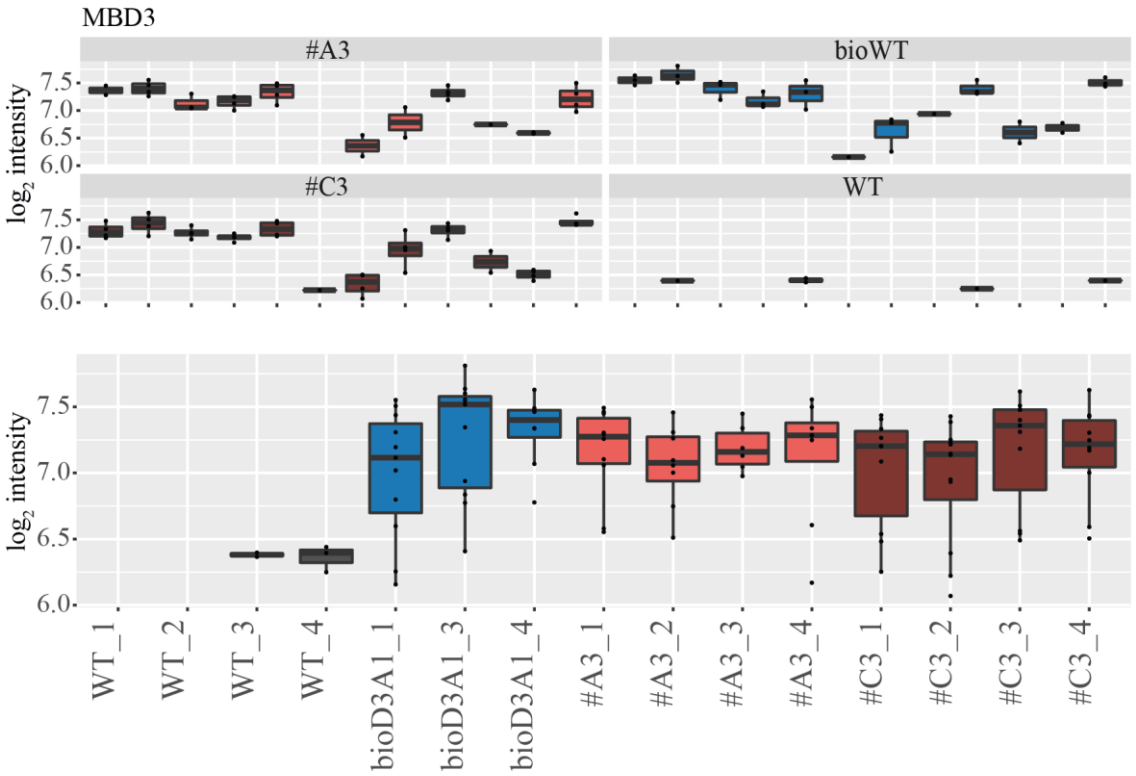


Figure 7.8. Peptide data for MBD3. Top panel shows individual peptide intensities per genotype group, where every box represents a unique peptide sequence along the protein; peptides are displayed in alphabetical order based on amino acid sequence. Bottom panel shows peptide intensity per individual sample. A single dot represents one detected peptide.

I also identified 14 proteins that were detected in both *bioDnmt3a*^{D329A} clone co-IPs but not in *Dnmt3a1* (Table 12). These proteins could be the new interactors of *Dnmt3a*^{D329A}. However, after a brief literature analysis and query of Online Mendelian Inheritance in Man database, barely any of the proteins are expected to localise in the nucleus and no link to DNMT3A could be made. Therefore, while I could confidently say that co-immunoprecipitation worked, the analysis of *bioDnmt3a1*^{D329A} proteins did not yield mechanistic insights into the aberrant methylation detected in adult tissues.

Table 12. Proteins uniquely co-immunoprecipitated in *bioDnmt3a1^{D329A} #A3 and #C3.*

Protein ID	**Mean intensity #A3**	**Mean intensity #C3**	***n* #A3**	***n* #C3**	**Genes**	**Protein names**
O88445 AURKC	20.95	22.35	2	4	*Aurkc Aie1 Aik3 Airk3 Ark3 Stk13*	Aurora kinase C (EC 2.7.11.1) (Aurora 3) (Aurora/IPL1-related kinase 3) (ARK-3) (Aurora-related kinase 3) (Aurora/IPL1/Eg2 protein 1) (Serine/threonine-protein kinase 13) (Serine/threonine-protein kinase aurora-C)
P11440 CDK1	22.29	23.18	1	1	*Cdk1 Cdc2 Cdc2a Cdkn1*	Cyclin-dependent kinase 1 (CDK1) (EC 2.7.11.22) (EC 2.7.11.23) (Cell division control protein 2 homolog) (Cell division protein kinase 1) (p34 protein kinase)
P68254 1433T	21.56	21.33	2	2	*Ywhaq*	14-3-3 protein theta (14-3-3 protein tau)
P84091 AP2M1	21.06	20.55	1	2	*Ap2m1 Clapm1*	AP-2 complex subunit mu (AP-2 mu chain) (Adaptor protein complex AP-2 subunit mu) (Adaptor-related protein complex 2 subunit mu) (Clathrin assembly protein complex 2 mu medium chain) (Clathrin coat assembly

						protein AP50) (Clathrin coat-associated protein AP50) (Mu2-adaptin) (Plasma membrane adaptor AP-2 50 kDa protein)
Q00547 HMMR	21.02	21.02	3	4	*Hmmr Ihabp Rhamm*	Hyaluronan mediated motility receptor (Intracellular hyaluronic acid-binding protein) (Receptor for hyaluronan-mediated motility) (CD antigen CD168)
Q3UGC7 EI3JA	22.13	19.60	1	1	*Eif3j1 Eif3s1-1*	Eukaryotic translation initiation factor 3 subunit J-A (eIF3j-A) (Eukaryotic translation initiation factor 3 subunit 1-A) (eIF-3-alpha-A) (eIF3 p35)
Q5SSZ5 TENS3	20.34	21.18	3	4	*Tns3 Tens1*	Tensin-3 (Tensin-like SH2 domain-containing protein 1)
Q80VY9 DHX33	21.03	22.45	1	2	*Dhx33*	ATP-dependent RNA helicase DHX33 (EC 3.6.4.13) (DEAH box protein 33)
Q8R2U0 SEH1	20.61	21.92	3	2	*Seh1l*	Nucleoporin SEH1 (GATOR complex protein SEH1) (Nup107-160 subcomplex subunit SEH1)
Q8R480 NUP85	20.59	21.57	3	3	*Nup85 Pcnt1*	Nuclear pore complex protein Nup85 (85 kDa nucleoporin) (FROUNT) (Nucleoporin Nup85) (Pericentrin-1)

Q9CYL5 GAPR1	20.69	21.12	2	2	*Glipr2* *Gapr1*	Golgi-associated plant pathogenesis-related protein 1 (GAPR-1) (Golgi-associated PR-1 protein) (Glioma pathogenesis-related protein 2) (GliPR 2)
Q9CZH8 CCD77	21.50	21.81	2	3	*Ccdc77*	Coiled-coil domain-containing protein 77
Q9JKF1 IQGA1	22.02	20.47	1	2	*Iqgap1*	Ras GTPase-activating-like protein IQGAP1
Q9WUA3 PFKAP	20.95	19.49	2	1	*Pfkp* *Pfkc*	ATP-dependent 6-phosphofructokinase, platelet type (ATP-PFK) (PFK-P) (EC 2.7.1.11) (6-phosphofructokinase type C) (Phosphofructo-1-kinase isozyme C) (PFK-C) (Phosphohexokinase)

7.3 *BioDnmt3a1*^{D329A} neuronal progenitor cell DNA methylation

Dnmt3a^{D329A} hypothalamus cells gain DNA methylation after the epigenetic programming event in early development, likely during differentiation and maturation of the neurons. To assess whether this pathology could be recapitulated in a cellular system, I differentiated *wild-type*, *bio-Dnmt3a1* and *bio-Dnmt3a1*^{D329A} ESCs to neuronal progenitor cells and assessed DNA methylation.

Six PBAT DNA methylation libraries were generated; two *wild-type* and two *bio-Dnmt3a1*, where two samples of the same genotype are technical replicates, and one each of clones #A3 and #C3 of *bio-Dnmt3a1*^{D329A}, where samples are biological replicates. The duplication levels were between 4.8-5.6% and samples had in the range of 33.1-44.9 million unique deduplicated reads per sample (Table 13).

Table 13. NPC PBAT library details.

Sample Name	Duplication (%)	Uniquely aligned reads (mln)
wt_1	4.8%	33.1
wt_2	5.4%	41.4
D3abio_1	5.6%	44.9
D3Abio_2	4.9%	35.8
bioD329A_#C3	5.3%	33.4
bioD329A_#A3	5.5%	40.9

Initial assessment of the samples showed a separation of *bio-Dnmt3a1* samples from *wild-type* and *bio-Dnmt3a1*^{D329A} clones (Figure 7.9). In agreement, it was observed that *bio-Dnmt3a1* cells had approx. 3% lower global DNA methylation (Figure 7.10). To understand the changes in the genome better, genomic tiles the size of 100 CpGs were

plotted for each pair of the genotypes. Unexpectedly, I discovered that *bio-Dnmt3a1* samples had a different distribution of DNA methylation than *wild-type*, and showed a cluster of tiles methylated much higher than the corresponding ones in the *wild-type* (Figure 7.11a). Even more surprisingly, *bio-Dnmt3a1*^{D329A} clones did not share this pattern; *bio-Dnmt3a1*^{D329A} methylation resembled *wild-type* (Figure 7.11b,c).

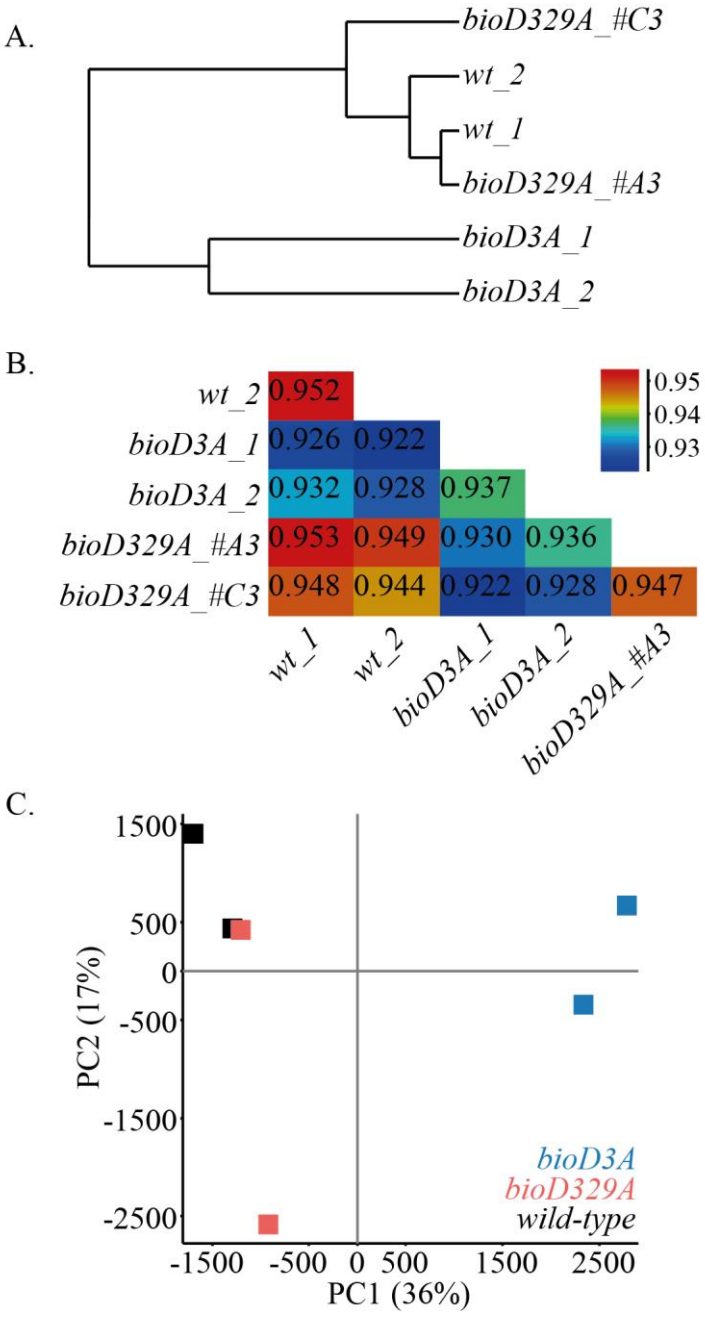


Figure 7.9. Separation of NPC DNA methylation samples. 100 CpG tiles were used in the analysis, individual samples are shown. **A.** A data similarity tree indicating Pearson correlation distances between individual samples. **B.** Pairwise correlation matrix for individual sample pairs, in scale 0 to 1. **C.** Principal component analysis.

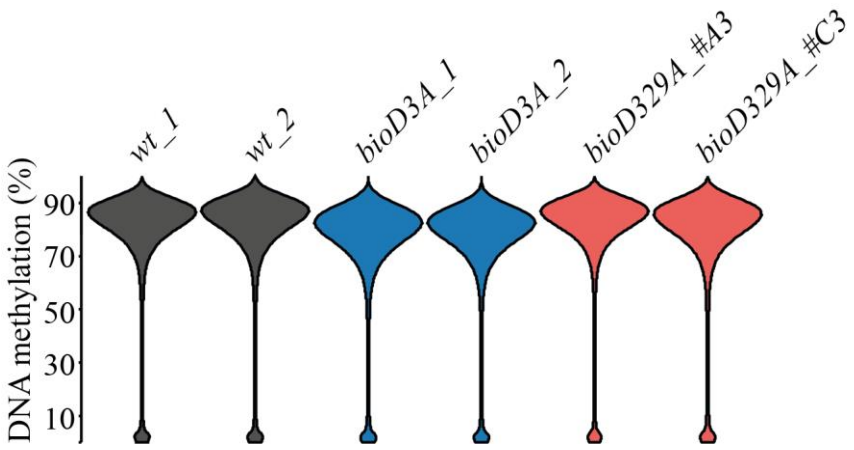


Figure 7.10. Global DNA methylation levels in NPCs. 100CpG tiles were used, each plot represents a single sample. Kruskal-Wallis rank sum test for genotype mean results in p-value < 2.2e-16.

To assess these discrepancies further, I first looked at different genomic features. Both *bio-Dnmt3a1* and *bio-Dnmt3a1*^{D329A} had higher DNA methylation levels than *wild-type* over CGIs and CGI promoters, although it was more pronounced in *bio-Dnmt3a1* cells (Figure 7.12a,b). Interestingly, statistical pairwise comparison found that differences between *bioDnmt3a1* and *bioDnmt3a*^{D329A} over CGI promoters were not statistically significant, suggesting that there is a convergence between these two genotypes at CGI promoters (Figure 7.12). *Wild-type* and *bio-Dnmt3a1*^{D329A} appeared to have the same level of DNA methylation over non-CGI promoters, gene bodies and intergenic regions (Figure 7.12b,c). Meanwhile, *bio-Dnmt3a1* was slightly hypomethylated in comparison, which would explain why on a global level it appeared to be methylated less than the other cell lines. Domains that become hypermethylated in the hypothalamus of adult *Dnmt3a*$^{\Delta/D329A}$ mice had approximately 23.6% DNA methylation in *wild-type* but 29.9% and 27.2% DNA methylation in *bio-Dnmt3a1* and *bio-Dnmt3a1*^{D329A}, respectively (Figure 7.13). Hypothalamus hypomethylated domains had 71.6%, 70.4% and 71.7% in *wild-type, bio-Dnmt3a1* and *bio-Dnmt3a1*^{D329A}, respectively.

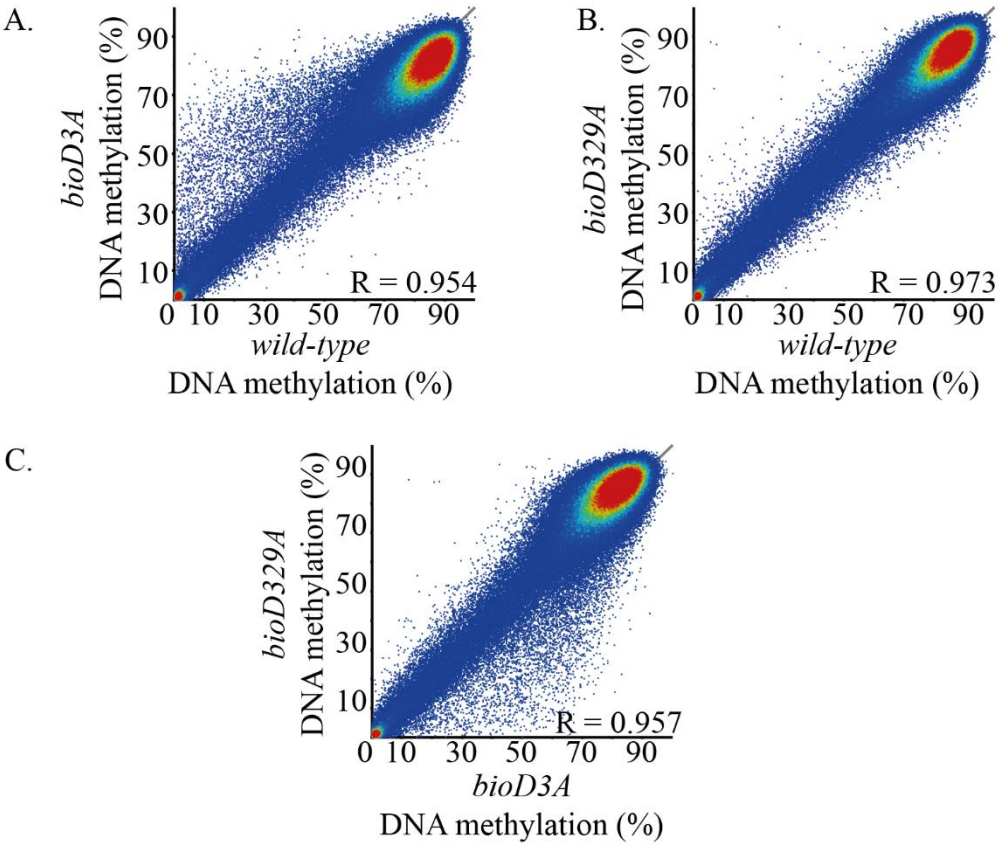


Figure 7.11. Assessment of NPC DNA methylation by 100 CpG tiles. Mean values per group are used in the analysis. Each dot represents 100CpG genomic tile.

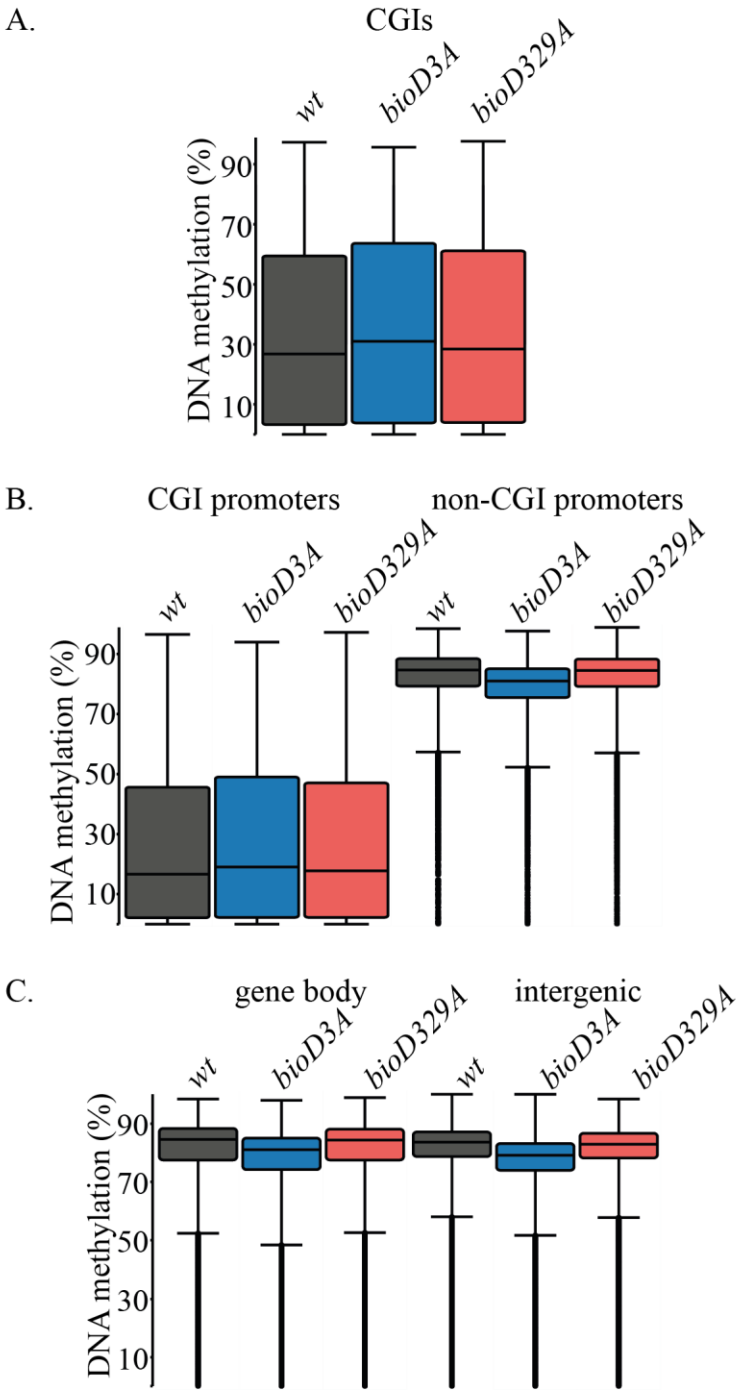


Figure 7.12. DNA methylation levels over different genomic regions. 100CpG tiles overlapping features indicated were used in the analysis, grouped per genotype. For **A.** all CGIs, **B.** non-CGI promoters, **C.** gene body and intergenic regions Kruskal-Wallis rank rum test p-value < 2.2e-16, and significant for all post-hoc pairwise Wilcoxon rank sum tests. The exception was **B.** CGI promoters, where Kruskal-Wallis p-value = 9.474e-10, but post-hoc Wilcoxon rank sum pairwise test for *bioDnmt3a1* against *bioDnmt3a1*^{D329A} p-value (Benjamin Hochberg adj.) = 0.056.

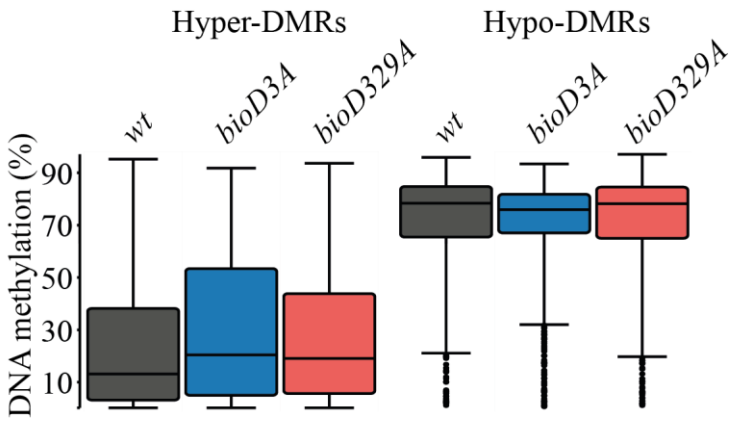


Figure 7.13. DNA methylation levels over hypothalamus DMRs. 100CpG tiles overlapping mouse hypothalamus DMRs were used in the analysis, grouped per genotype. For hyper-DMRs, Kruskal-Wallis rank rum test p-value was < 2.2e-16, and significant for all post-hoc pairwise Wilcoxon rank sum tests. For hypo-DMRs, Kruskal-Wallis rank rum test results in p-value = 3.453e-07, but post-hoc Wilcoxon rank sum pairwise test for *wild-type* against *bioDnmt3a1*^{D329A} p-value (Benjamin Hochberg adj.) = 0.54.

Initially, I expected to use the *bioDnmt3a1* cell line for DNA methylation comparison with *bioDnmt3a1*^{D329A}, given that it would have the same level of expression of the non-endogenous *Dnmt3a1* and the only difference between the two cell lines would be the presence of the D329A mutation. However, global and feature-based DNA methylation differences suggested that these cell lines are more diverse than anticipated; further pairwise comparisons were conducted to understand these differences.

First, the edgeR statistical tool, with *p*-value <0.01 and absolute difference >10% cut-off, was used to identify differentially methylated tiles between *wild-type* and *bioDnmt3a1*. There were 4,124 hypermethylated and 5,275 hypomethylated DMRs. Only 0.3% hypermethylated tiles matched hypothalamus hypo-DMRs, and 20% of hypermethylated tiles matched adult *Dnmt3a*$^{\Delta/D329A}$ hypothalamus hyper-DMRs. Approx. 87.8% of hypermethylated tiles overlapped CGIs. Gene ontology analysis indicated two major trends for genes that were surrounded by hypermethylated domains: transcription factor and neuronal function-related genes (Figure 7.14). Hypomethylated domains were enriched in genes associated with neuronal function and neuronal cellular compartments, such as synapse (Figure 7.15).

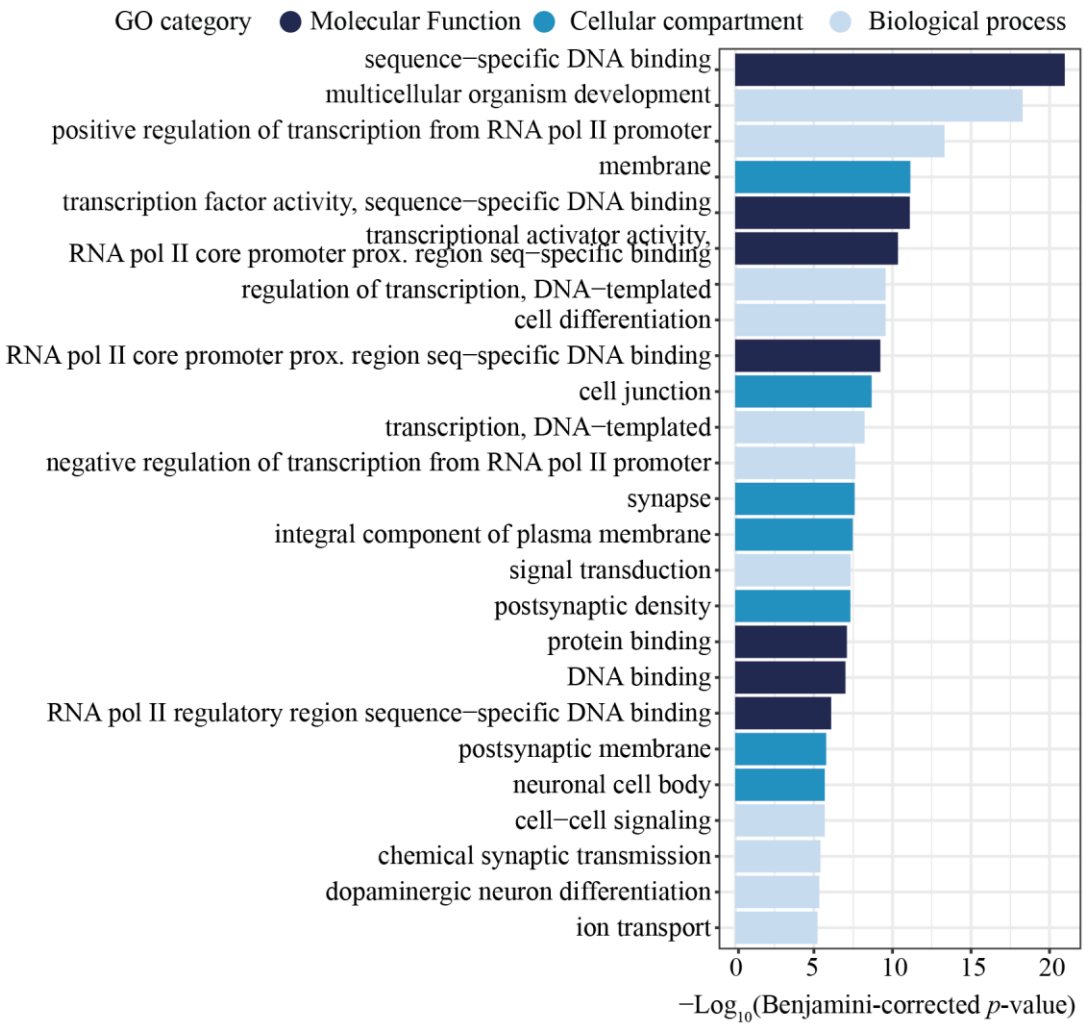


Figure 7.14. GO analysis of genes linked to domains hypermethylated in *bioDnmt3a1* NPCs compared to *wild-type*.

Next, I looked at differentially methylated tiles between *wild-type* and *bioDnmt3a1*^{D329A}. There were only 232 DMRs, of which 198 were hypermethylated and 34 were hypomethylated. Approx. 57.6% of those hypermethylated were shared with adult *Dnmt3a*$^{\Delta/D329A}$ hypothalamus hyper-DMRs; while none of the hypomethylated domains matched hypothalamus hypo-DMRs. Gene ontology analysis of genes within 2kb distance to DMRs indicated that hypermethylation sites were enriched for transcription factor and developmental genes (Figure 7.16). Meanwhile, there were only 19 genes nearby the hypomethylated tiles, and no significant enrichment could be identified.

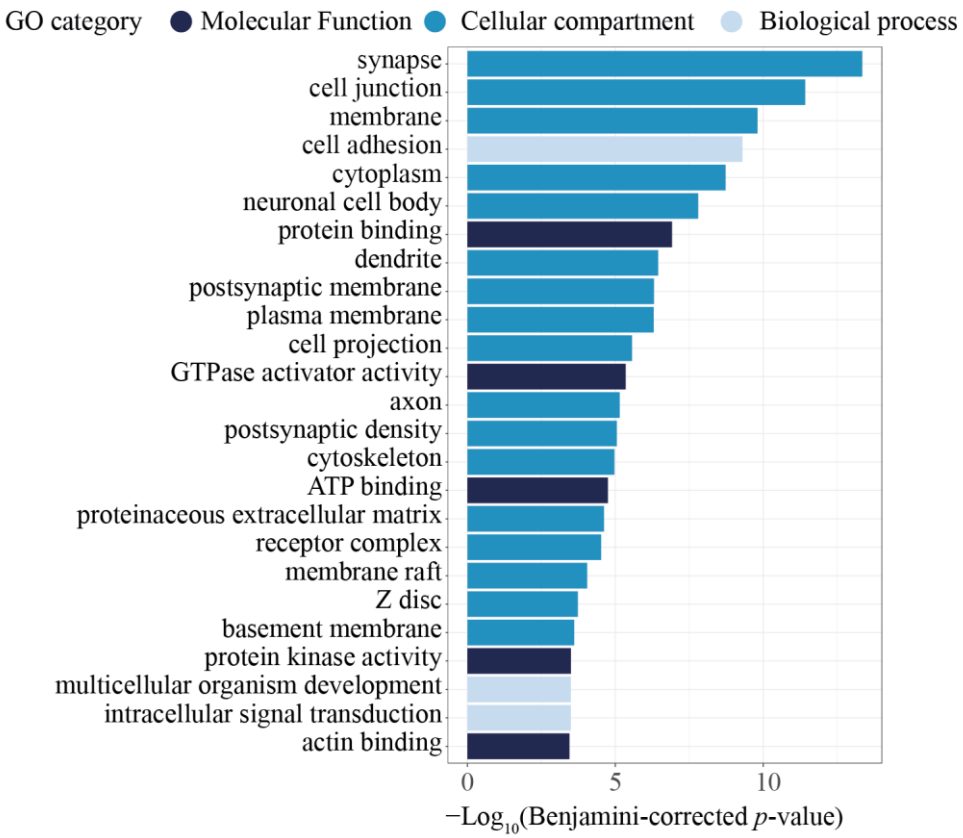


Figure 7.15. GO analysis of genes linked to domains hypomethylated in *bioDnmt3a1* NPCs compared to *wild-type*.

Finally, a comparison of *bioDnmt3a1* and *bioDnmt3a1*^{D329A} cells found 2,756 hyper- and 2,688 hypomethylated DMRs, of which only 2% and 0.6%, respectively, matched hypothalamus DMRs. Gene ontology analysis of hypermethylated domains showed enrichment in protein binding and cellular compartments of cytoplasm, membrane and synapse (Figure 7.17). Hypomethylated domains were enriched for a mixture of transcription related functions but also membrane, cell junction, and neuronal cell body compartments (Figure 7.18).

Together, this analysis suggested complex dynamics. It would appear that the "positive control" *bioDnmt3a1* cell line was rather divergent, and could not be used as a meaningful control in this case. Meanwhile, *wild-type* and *bioDnmt3a1*^{D329A} cell lines were very similar but had a small number of DMRs. Given that the majority of the DMRs exhibited gain of DNA methylation in the presence of the mutation, and that half of those would be hypermethylated in adult mouse *Dnmt3a*$^{\Delta/D329A}$ hypothalamus, I suggest that these cells could be at the start of a trajectory observed in hypothalamus. Since there are only about 200 differentially methylated sites, neuronal progenitor cells are not an ideal model, and

further differentiation and maturation into neurons could allow for greater accumulation of change.

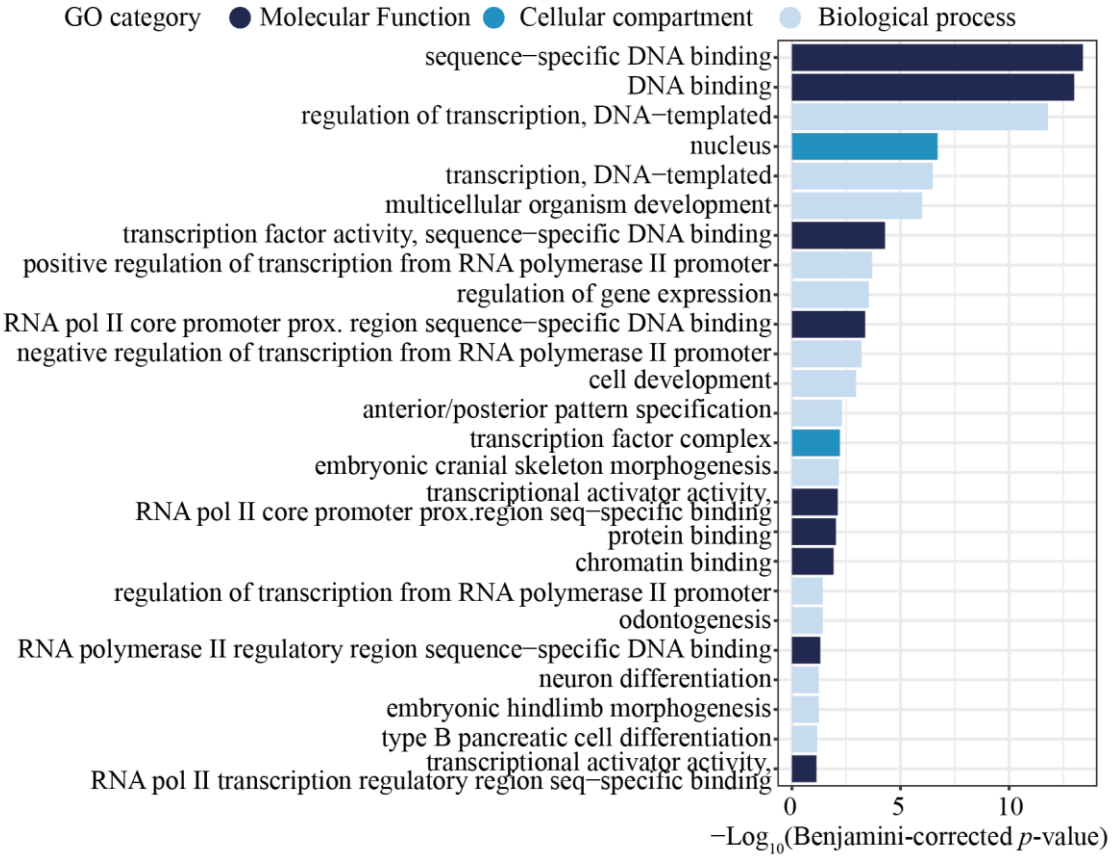


Figure 7.16. GO analysis of genes linked to domains hypermethylated in *bioDnmt3a^{D329A}* NPCs compared to *wild-type*.

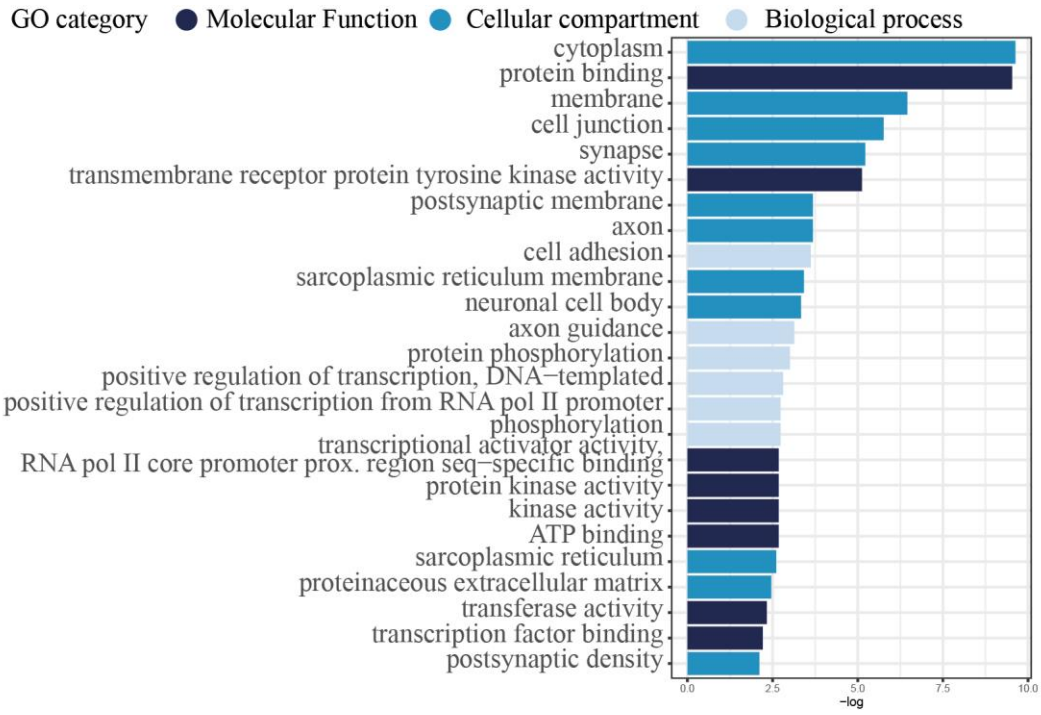


Figure 7.17. GO analysis of genes linked to domains hypermethylated in *bioDnmt3a1*^{D329A} NPCs compared to *bioDnmt3a1*.

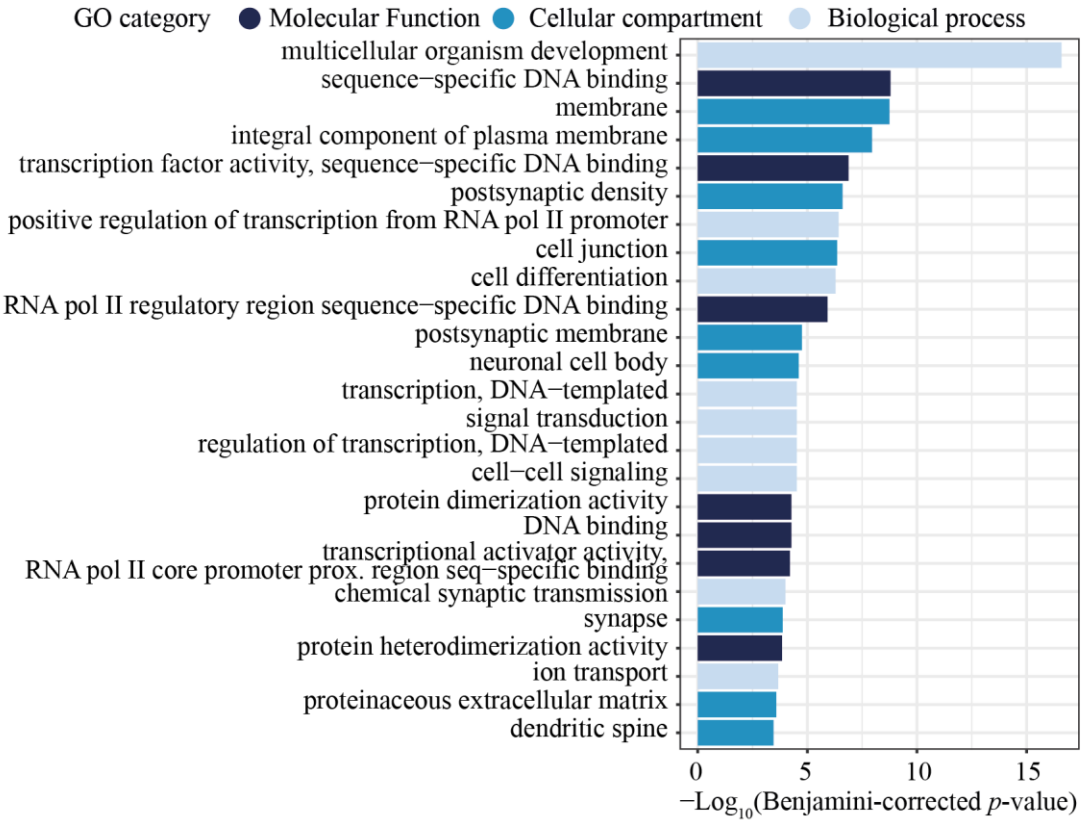


Figure 7.18. GO analysis of genes linked to domains hypomethylated in *bioDnmt3a1*^{D329A} NPCs compared to *bioDnmt3a1*.

7.4 Summary and discussion

In this chapter, I show successful generation of ESCs carrying *bioDnmt3a1*^{D329A}. Unfortunately, no novel interactors of DNMT3A could be found and the gain of aberrant DNA methylation in neuronal progenitor cells was rather limited.

The aim of work described in this chapter was to overcome the complexity of brain tissue cell type heterogeneity and simplify biochemical experimental approaches by using the RAMBiO system (Baubec et al., 2013). Given the knowledge from previous chapters, it was expected that mutant DNMT3A would have an effect on DNA methylation even in the presence of the *wild-type* allele. Thus, a *wild-type* background was used for experiments described here.

This design has its limitations: ESCs already express both *de novo* DNMTs 3A and 3B to a relatively high level and both of these play a role in ESC DNA methylation establishment (Chen et al., 2003; Gu et al., 2018). Biotinylated DNMT3A has to compete with endogenously expressed protein and free-floating wild-type DNMT 3A, 3B, and 3L might be tethering the mutant protein (Chédin et al., 2002; Chen et al., 2005; Suetake et al., 2004). This hypothesis is supported by my Co-IP-MS experiment result, where all three DNMT3s are identified as significant and some of the most common interactors. Expression of the biotinylated mutant protein in *Dnmt3a/Dnmt3b* double knock-out (DKO) cells would eliminate this possibility and the targeting of bioDNMT3A^{D329A} could be assessed. But the fact that cell line with DNMT3A^{D329A} does not show lower DNA methylation than *wild-type* suggest that it does not instigate a dominant negative effect.

It is plausible that the reported reduction in DNMT3A-D329A association with chromatin (Dhayalan et al., 2010) in a system where expression and replication levels are high would result in the mutant protein not being able to linger around long enough to cause the methylation gain. The requirement of low replication rate could be solved if neuronal progenitors were matured in culture. While ESCs survive without DNA methylation, they fail to differentiate successfully, and in this case *wild-type* background would be required (Chen et al., 2003).

There were no changes observed in DNA methylation of E7.5 embryos, therefore I hypothesised that it is unlikely to see any DNA methylation changes in serum-grown ESCs, which would have been originally derived from a similarly aged embryo (Zwaka and Thomson, 2005), and immediately started with NPCs for assessment of DNA

methylation. However, CoIP-MS requires high numbers of cells and replicates to generate robust results and usage of ESCs instead of NPCs was an experimental compromise. If the assumption is that aberrant DNA methylation cannot be established in ESCs because of a high replication rates, it does not necessarily imply that the potential novel interactor is not present and/or bound to DNMT3A in the ESC environment. Similarly, *AviTag* is a very small tag where biotin is added as a post-translation modification (Fairhead and Howarth, 2015), hence it is unlikely to disrupt many protein-protein interactions.

Assessment of DNA methylation in NPCs carrying *bioDNMT3A1*^{D329A} found a very limited effect on DNA methylation compared to *wild-type*, but why the *bioDNMT3A1* cell line had many differentially methylated sites is rather unclear. During the transition from ESCs to NPCs lineage specificity is acquired by changes in DNA methylation and chromatin environment (Mohn et al., 2008). The CoIP-MS experiment does not indicate any difference between *bioDNMT3A1* and *bioDNMT3A1*^{D329A} cell line proteomes, at least at the ESCs stage, but does not give any insight into differences arising in NPCs.

The *bioDnmt3a1* cell line has been generated independently and kept in culture for multiple passages, and therefore could have accumulated genetic and epigenetic alterations, as observed in human ESCs (Tanasijevic et al., 2009). A better control would have been introduction of *bioDNMT3A1* into reference *wild-type* HA36CB1 ESC cell line at the same time as *bioDNMT3A1*^{D329A}, which would provide confidence that differences observed are due to the introduced transcripts rather than culture conditions and handling.

Overexpression of *Dnmt3b* in adult mouse tissues resulted in slightly increased levels of DNA methylation over CGI-rich domains in all cells and tissues (Zhang et al., 2018). Assuming that accumulation of aberrant DNA methylation could occur gradually in the cells with increased levels of DNMT3A, it is possible that the *bioDnmt3a1* line had accumulated changes and, hence, why I see a gain in DNA methylation. Similarly, as in the *Dnmt3b* overexpression study, here CGI-rich regions are those showing higher levels of methylation.

While protein epitope tagging in simplified systems, such as cell culture, is a great way to study them, the results here highlight the challenges and failure to recapitulate the effect a mutation can have in a complex organism.

8 Discussion

The aim of the work described in this thesis was to investigate the role of the PWWP domain in DNMT3A in targeting to the genome. To that end, I used a mouse model carrying the D329A mutation in the aromatic cage of the PWWP domain (Sendžikaitė et al., 2019). This mutation has been previously described to ablate the interaction between the DNMT3A PWWP domain and H3K36me3 *in vitro* and was considered as a robust loss-of-function mutation, used as a control in other biochemical work (Bock et al., 2011; Dhayalan et al., 2010; Dukatz et al., 2019; Kungulovski et al., 2014; Mauser et al., 2017).

In characterising this mouse, I found that the D329A mutation does not lead to loss of DNA methylation over gene bodies, marked by H3K36me3, which I confirmed in different tissues and time points. However, I identified that it is a gain-of-function mutation, which causes postnatal growth retardation in mice and progressive accumulation of DNA methylation over H3K37me3-marked and bivalent chromatin. This hypermethylation results in altered chromatin landscape and derepression of some developmental genes. Subsequently, I established an ESC line where biotinylated mutant DNMT3A is expressed. Unfortunately, I could not identify any novel binding partners of the bioDNMT3A^{D329A} protein in ESCs and saw rather limited changes in DNA methylation in NPCs, but this cell line can be used in further work.

8.1 The PWWP domain and H3K36me2/3

Given a tight link between gene expression, H3K36me3, and DNA methylation in the oocyte, the initial hypothesis appeared very well grounded and the oocyte system an ideal platform to test it (Veselovska et al., 2015; Xu et al., 2019). Therefore, it was a great surprise to see a lack of strong loss of DNA methylation over H3K36me3 domains.

The PWWP domain also recognises H3K36me2 (Dhayalan et al., 2010; Dukatz et al., 2019). Recently, H3K36me2 was found to specifically recruit DNMT3A to non-coding euchromatin regions via the PWWP domain and to stimulate its catalytic activity in mouse mesenchymal and human breast cancer cell lines (Weinberg et al., 2019; Xu et al., 2020). These studies reported that loss of H3K36me2 led to a significant relocalisation of

DNMT3A to gene bodies. Upon introduction of a D333A mutation, which is only a few amino acids downstream from D329A, this relocalisation was lost. Based on interpretation of the accompanying ChIP-qPCR data, there is no significant loss of DNMT3A-D333A binding over H3K36me2 domains compared to wild-type DNMT3A, and changes at bivalent chromatin are not assessed (Weinberg et al., 2019).

In the context of the Weinberg et al. (2019) study, it could be hypothesised that H3K36me2/3 acts as a tethering mark, and DNMT3A is recruited to H3K36me2 intergenic chromatin and, to a lesser extent, H3K36me3 at gene bodies. Since DNMT3B is thought as the major activity methylating gene bodies (Baubec et al., 2015), it can be assumed DNMT3A is only localising to those regions transiently or due to the overlap between H3K36me2 and K36me3; but it does not have any functional effect over gene bodies. If, upon introduction of a mutation in the PWWP domain, H3K36me3 binding is lost, and interaction with H3K36me2 is only slightly affected, the mutant DNMT3A would increasingly relocalise to the place of next affinity – bivalent chromatin shores (Gu et al., 2018; Jeong et al., 2014; Manzo et al., 2017).

This theory is somewhat compatible with my data, where I see a small loss of DNA methylation over intergenic domains. Compared to the distribution of random genomic tiles, CGI-overlapping intergenic tiles are overrepresented (Figure 5.25). Yet, CGI-containing intergenic tiles constitute only 43 out of 600 (7.2%) tiles that are hypomethylated in adult *Dnmt3a*$^{\Delta/D329A}$ hypothalamus, suggesting that recruitment of DNMT3A^{D329A} to intergenic domains is not affected strongly. The model proposed would rely on high abundance of DNMT3A in the cell together with long residence time on chromatin, so that the strength of enzyme-chromatin interaction would in proxy control the relative binding of DNMT3A to different targets. It would also require DNMT3A to still be recruited and/or stay resident over a H3K36me2/3 domains once the DNA is methylated at those loci.

What is intriguing though, is that the oocyte chromatin landscape is completely different, and H3K36me3 is enriched at sites of DNA methylation (Stewart et al., 2015; Xu et al., 2019). Deletion of SETD2, an enzyme catalysing H3K36me3, leads to loss of DNA methylation over H3K36me3 domains, which are normally hypermethylated in the oocyte, and low but widespread gain at hypomethylated domains (Xu et al., 2019). The H3K36me2 was found to overlap H3K36me3 and is only associated with transcribed gene bodies, therefore the DNA methylation is SETD2 and H3K36me3 dependent (Shirane et

al., 2020). This evidence really suggests that H3K36me3 and DNA methylation are very strongly linked in the oocyte. Since I do not see change in the oocyte DNA methylation over H3K36me3 domains, it can be hypothesised that the D329A mutation fails to disrupt the link between H3K36me3 and the PWWP domain in the oocyte. Does the mutation fail to interrupt the interaction or is the PWWP domain not involved at all? Given that in the oocyte DNMT3A is in a complex with DNMT3L, it could be that DNMT3L recruitment acts as a dominant mechanism for genomic targeting. Alternatively, the D329A might be a hypomorphic mutation but the effects are masked by an extensive timeframe of oocyte growth and *de novo* methylation establishment.

Together with findings of other studies, I suggest that the function of DNMT3A *in vivo* may be associated with H3K36me2/3, but that the D329A mutant protein does not affect this dynamic to an extent where functional change is observed.

8.2 The two DNMT3A isoforms

I chose two main tissues where DNMT3A is the dominant DNA methyltransferase: the mature oocyte and the adult brain (Feng et al., 2005; Kaneda et al., 2010, 2004), so that redundancy with DNMT3B could be avoided. The system where DNMT3B was not expressed, like the adult brain, or was shown to be non-functional and not to rescue DNA methylation, like the oocyte, was of particular importance when studying a domain shared between the two *de novo* methyltransferases.

Although the original study by Dhayalan et al. (2010) tested the shorter DNMT3A2 isoform, my initial results from oocytes, where only DNMT3A2 is present, did not show any change. Discovery of the postnatal onset, growth restriction phenotype, which time-wise overlaps the switch from DNMT3A2 to DNMT3A1 expression (Manzo et al., 2017), provoked an idea that N-terminal domain might be involved in promoting the observed ectopic methylation. The PWWP domain is virtually the same between the two methyltransferases, while the N-terminal is different, and findings in ESCs showed that only the DNMT3B-PWWP domain was required for H3K36me3-marked gene body methylation (Baubec et al., 2015). If different domains can act in concert to target DNMTs, such as auto-inhibition of the catalytic domain by the ADD domain (Guo et al., 2015), it is feasible that the less well-studied intrinsically-disordered N-terminal domain could have an impact on DNMT3A localisation through another domain, such as the PWWP.

The long N-term containing DNMT3A1 isoform is expressed from late embryogenesis into adulthood in mice (Manzo et al., 2017). Work by Manzo et al. (2017) identified that the N-terminal domain was essential in DNMT3A localisation to the shores of bivalent chromatin. A similar study found that both DNMT3A1 and DNMT3A2 localise to the shores of DNA methylation canyons in ESCs (Gu et al., 2018). In the light of these findings, I initially hypothesised that the longer isoform would be the one responsible for establishment of hypermethylated H3K27me3 domains in *Dnmt3a*^{D329A} mice. However, mutant DNMT3A1 was not sufficient to introduce dramatic gain in DNA methylation in NPCs. Therefore, I can conclude that the isoform alone is not the definitive factor by which the DNMT3A PWWP domain is recruited to the genome.

8.3 Aberrant DNA methylation and H3K27me3

Unlike the strong antagonistic relationship of DNA methylation and H3K4me3 (Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010), the link between methylation and H3K27me3 is complex and not understood so well. While they are mutually exclusive at CGIs, there is a certain degree of plasticity where both DNA methylation and H3K27me3 can co-exist (Brinkman et al., 2012; Statham et al., 2012), as observed in this study.

I found that H3K27me3-marked chromatin specifically gained aberrant DNA methylation, but *in vitro* studies did not see the mutant PWWP domain acquiring affinity to this PTM and there is little evidence of direct DNMT3A-H3K27me3 interaction (Dhayalan et al., 2010; Dukatz et al., 2019; Manzo et al., 2017; Mauser et al., 2017). While DNMT3s have a strong affinity towards H3K36me2/3, there may be a weak affinity towards H3K27me3, such that the abundance of different DNMT3s and their interacting proteins could modulate chromatin localisation depending on the cellular context.

Recently, point mutations in the PWWP domain of human *DNMT3A* have been described in a specific class of tumour – hereditary paraganglioma (Mellid et al., 2020; Remacha et al., 2018) and in microcephalic dwarfism (Heyn et al., 2019), including a substitution in the amino acid residue orthologous to murine D329. In both cases, the mutations are associated with gain of methylation of H3K27me3-marked domains containing developmental regulatory genes, similar to the findings in my study.

Heyn et al. (2019) had suggested non-selective relocalisation of mutant DNMT3A from H3K36me2/3 to H3K27me3 sites, simply because the latter are unmethylated. However, my detailed analysis of the ectopic DNA methylation in relation to chromatin data suggests a role of positive recruitment of the DNMT3A-PWWP mutant protein to H3K27me3-marked domains, as I do not observe gain of DNA methylation across other unmethylated genome features. It is possible that a conformation change caused by the D329A mutation somehow enhances the affinity between DNMT3A and H3K27me3, but other proteins only present in an *in vivo* environment are likely to contribute to this interaction.

A direct association of DNMTs with H3K27me3 regions in some contexts is supported by the observation that DNA methylation is dependent on the deposition of H3K27me3 at a subset of CGIs in extraembryonic ectoderm (Smith et al., 2017). Similarly, H3K27me3 sites in ESCs often gain aberrant DNA methylation in cancer (Ohm et al., 2007; Schlesinger et al., 2007; Widschwendter et al., 2007). In cancer, this hypermethylation leads to silencing of differentiation genes and, therefore, to an enhancement of self-renewal. In contrast, during development, DNMT3A can silence polycomb-target genes in a tissue-specific manner, therefore promoting differentiation and lineage determination (Yagi et al., 2020). And although both DNA methylation and H3K27me3 are considered repressive modifications, *Hox* genes can be activated when DNA methylation is targeted to their gene bodies in tumours (Su et al., 2018). This is similar to my observation of de-repression of strongly silenced genes in $Dnmt3a^{\Delta/D329A}$ hypothalamus, where methylation is acquired and H3K27me3 is maintained. Moreover, theoretically, DNA methylation gain could cause alternative exon usage (Gelfman et al., 2013; Shayevitch et al., 2018). These examples highlight how functional outcomes of DNA methylation gain are very context-dependent and challenging to decipher.

H3K27me3 is catalysed by the PRC2 (Cao et al., 2002), while the PRC2 protein EZH2 has been shown to recruit DNA methylation to its target promoters (Viré et al., 2006). DNMT3A has also been found to interact with PRC2 when establishing the boundaries between the DNA methylated and PRC2-repressed chromatin (Gu et al., 2018; Manzo et al., 2017). The regulation of these hypomethylated, H3K27me3-marked domains involves a complex and still poorly understood interplay between PRC2, DNMT3A and TET DNA demethylases (Figure 8.1). During lineage commitment, DNMT3A can methylate gene bodies of PRC2 targets in a tissue-specific manner (Yagi et al., 2020),

which suggests the mechanisms of permitting DNMT3A to enter the core of these domains are in place. DNA methylation generally, albeit not always, inhibits PRC2 binding and activity *in vitro* (Bartke et al., 2010; Wang et al., 2017) and can block the deposition of H3K27me3 (Brinkman et al., 2012; Jermann et al., 2014). In this study, I found H3K27me3, albeit with a tendency to be reduced, to remain at the hypermethylated DMRs. This can be due to the non-proliferative environment in the neurons, where nucleosomal turnover is relatively low or simply be a signal coming from different cells, as hypothalamus is such a heterogeneous tissue (Chen et al., 2017).

I would speculate that DNMT3A regulation at PRC2-bound bivalent chromatin could be altered by the D329A mutation. At this time, it is unclear how the D329A mutation may have altered the properties of DNMT3A to enhance its interaction with H3K27me3 and/or with members of the PRC2, allowing it not only to target the boundaries but the protected core of bivalent domains as well as (Figure 8.1b). One possibility is that H3K36me2/3 acts as a sink for wild-type DNMT3A, and loss of affinity for this interaction would permit mutant protein to localise specifically to bivalent chromatin, therefore overloading DNMT3A at the bivalent chromatin shores.

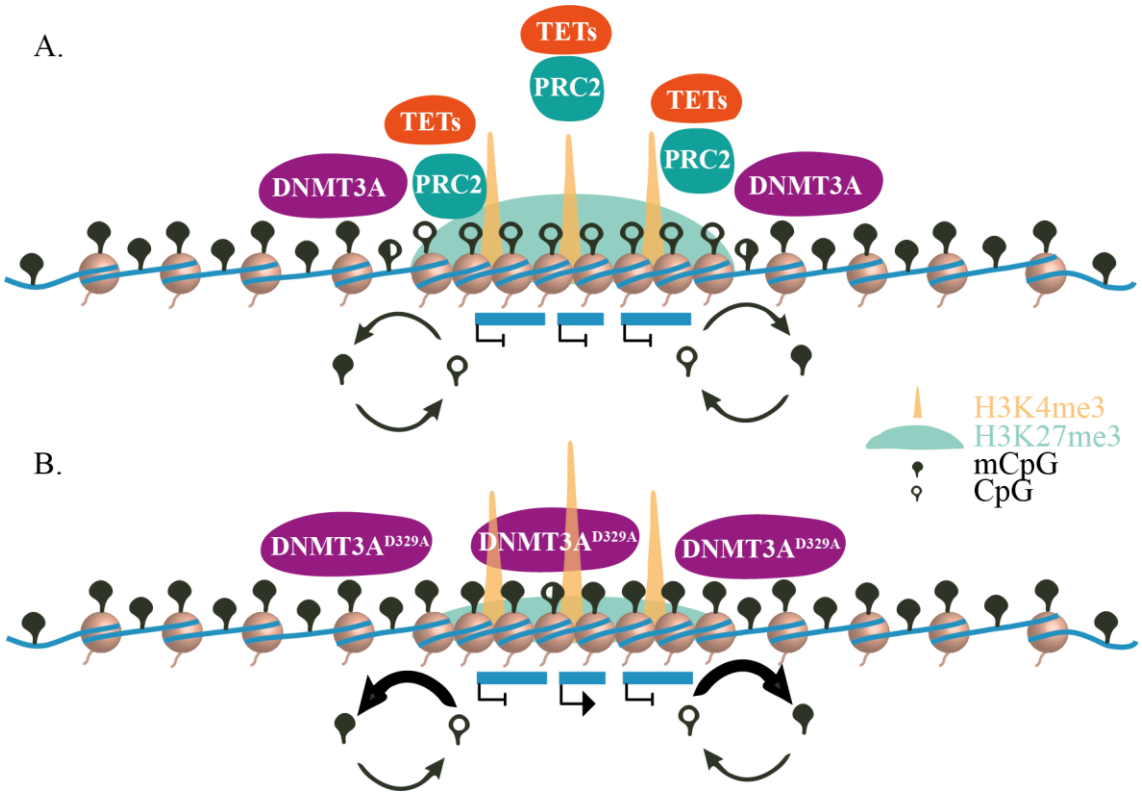


Figure 8.1. Model of molecular mechanisms in act at bivalent domains. A. A representation of current understanding of bivalent domain regulation. The PRC2 complex deposits H3K27me3 and compacts chromatin across the promoters and surrounding regions of transcriptionally silent, developmental genes (Boyer et al., 2006). TET proteins are recruited, which result in active removal of DNA methylation and protection of bivalent chromatin from *de novo* DNMTs (Gu et al., 2018). DNMT3A is recruited to bivalent chromatin shores through an unknown mechanism, enabling active establishment of DNA methylation at the boundaries (Gu et al., 2018; Jeong et al., 2014; Manzo et al., 2017). Through the competing actions of TETs and DNMT3A there is a cycle of DNA methylation turnover at the boundaries of bivalent domains. **B.** Representation of bivalent chromatin dynamics in the presence of DNMT3A^{D329A}. The mutant protein is able to access bivalent chromatin and establish methylation across the whole domain. The presence of DNA methylation appears to influence the local chromatin environment and result in H3K27me3 reduction, potentially due to exclusion of PRC2 from methylated DNA. As a consequence, some genes become de-repressed and show an increase in promoter H3K4me3, while others maintain their transcriptional silencing; this likely dependents on the availability and strength of necessary transcription factors.

There has been considerable focus on what defines the boundaries of these domains, which is thought to rely on a dynamic equilibrium between DNMT3A and TET proteins (Gu et al., 2018; Jeong et al., 2014; Li et al., 2018; Manzo et al., 2017). Knock-outs of either protein strongly affect domain boundaries, whilst the core regions remain protected through an unknown mechanism. The protection of these domains is exemplified by a *Dnmt3b* overexpression mouse model that shows aberrant gain of methylation genome-wide and a marginal increase (<10%) across H3K27me3-marked domains (Zhang et al., 2018). Therefore, it is likely that relocalisation of irregularly high levels of mutant DNMT3A could tip the balance between DNMT3A and TETs at those sites, and they are no longer protected from DNA methylation (Figure 8.1). Together, these findings suggest that aberrant targeting of DNA methylation could occur not solely due to direct H3K27me3 recruitment, but due to changes in protein-protein interactions, in particular with PRC2 or quantitative imbalance with TETs.

8.4 Future directions and concluding remarks

Work presented in this thesis provides insight into the complex ways of how DNMTs use their domains in order to faithfully establish DNA methylation, and how they are guided by chromatin PTMs. Many aspects of the mechanisms remain unclear and will benefit by further investigations.

First, I would like to provide general notes, which should be taken in consideration when designing subsequent studies of DNMTs. Whenever a new model is designed, whether it is a mouse or a cell, it is important to introduce small protein tags at endogenous DNMT sites. Ideally, these should be tested in *wild-type* genetic background to ensure there is no interference of protein-protein interactions. These tags could be different for the two DNMT3A isoforms, which would allow clarification whether a particular one is involved in the process of interest. Multiple laboratories have successfully used these tags and there appear to be no significant drawbacks (Emperle et al., 2018; Fuks et al., 2001; Gao et al., 2020; Manzo et al., 2017; Weinberg et al., 2019), while they provide easier immunoprecipitation, staining, and blotting for assessment of presence of DNMTs.

Studies of DNMTs can benefit greatly by use of systems such as oocyte and the brain. For the investigation of H3K36me3 recruitment of DNMT3A specifically, it would be very helpful to generate a model where DNMT3B expression is definitely silenced and/or the PWWP domain of DNMT3A is deleted. This should provide insights whether the

domain is really needed for H3K36me3 recruitment and clarify if the D329A mutation has a different outcome *in vivo*.

For the investigations in the brain, constructs missing PWWP and/or N-terminal domain could be used. The N-terminal domain target DNMT3A1 to bivalent chromatin shores but is it the *wild-type* PWWP domain responsible for autoinhibition at the bivalent chromatin core? Does ablation of the whole domain recapitulate the D329A mutation and allow targeting of aberrant DNA methylation to canyons? My findings that gain-of-function effects of the DNMT3A-D329A mutant cannot be easily recapitulated in neuronal progenitor cells suggest that these defects emerge or accumulate later in development. Use of cerebral organoids, which recapitulate embryonic brain development and can be matured in a dish (Marshall and Mason, 2019), would allow the scale and assessment of multiple constructs, and allow introduction of combinations of mutations, such as knock-out of H3K36me2/3 and H3K27me3 catalysing enzymes. Also, this would reduce the resources, time and use of animal models. Similarly, organoids can be built using human ESCs, and therefore DNMT3A mutations causing Tatton-Brown-Rahman overgrowth or microcephalic dwarfism could be studied specifically in an environment resembling human brain (Heyn et al., 2019; Tatton-Brown et al., 2014).

The *Dnmt3a-D329A* mouse model still has a lot of potential to answer questions of how DNMT3A is involved in physiology and growth regulation. Further assessment of brain histology would allow elucidation of microcephalic tendency. Is there a defect in neuronal proliferation or differentiation? Is the migration and spreading of the neurons affected in the brain? Which cell types fail to be established or are overrepresented in the mutant brain? Finding the path from mutation to molecular pathology to the phenotype is the key to understanding a vast number of developmental disorders linked to epigenetic regulators, and this model can be used to further investigate microcephalic dwarfism (Heyn et al., 2019). Similarly, the biotinylated-DNMT3A-D329A mutant cell line I established could be used to investigate the chromatin binding of mutant DNMT3A. It is a remaining question whether the whole protein relocalise, or is it is generally bound to the affected areas but the chemical methylation reaction can only proceed when some safeguarding mechanisms are down. The question remains whether in cancer environment, where hypermethylation is found, mutations have inactivated safeguarding mechanisms, and whether similar mechanisms are in play in the presence of the D329A

mutation. Similarly, the biotinylated-DNMT3A-D329A mutant cell could be used for neuronal maturation and in organoid formation.

For investigation of the impact of DNA methylation on the local chromatin environment, CRISPR-Cas Sun-Tag recruitment of DNMT3A could be used (Huang et al., 2017). This method employs recruitment of protein complex to genomic loci using catalytically inactive Cas9 with a guide. Cas9 is fused with antibody epitopes. The protein of interest, e.g. DNMT3A, is then also fused to single-chain variable fragment (scFv) which then guides it to the target locus for DNA methylation. A less specific approach is binding DNMT3A catalytic domain to zinc-fingers with promiscuous affinity to chromatin (Broche et al., 2020). This way a hypothesis of non-targeted DNA methylation relocalisation could be assessed. These and similar methods could be used to methylate different genes which gain DNA methylation in *Dnmt3a*^{D329A} mouse, and quantify the contribution of this to underlying gene expression, local chromatin environment and the growth phenotype. Another advantage of this system would be the ability to induce these changes using temporal- and tissue-specific recombinases and to identify critical timepoints in development.

In summary, here I provide a characterisation of the growth phenotype and underlying molecular changes of the *Dnmt3a*^{D329A} mutant mouse model. I found that the interaction between H3K36me3 and the mutant DNMT3A is not altered in the oocyte. I demonstrate that the mutant DNMT3A exhibits specific, altered targeting to H3K27me3-marked chromatin in the mouse brain. The aberrant gain in DNA methylation across these domains is only observed in postnatal development, suggesting that the non-proliferative environment, coinciding with DNMT3A1 isoform expression in adult tissues, unmasks this effect. This study provides further insights into the regulation of bivalent chromatin domains and the role of the DNMT3A PWWP domain. This model will be valuable for future investigations into the role of DNA methylation abnormalities in cancer and the growth regulatory function of DNMT3A.

9 References

Aapola, U., Kawasaki, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K., Antonarakis, S.E., Shimizu, N., Kudoh, J., Peterson, P., 2000. Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 65, 293–298. https://doi.org/10.1006/geno.2000.6168

Auclair, G., Borgel, J., Sanz, L.A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné, T., Feil, R., Weber, M., 2016. EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. Genome Res. 26, 192–202. https://doi.org/10.1101/gr.198291.115

Auclair, G., Guibert, S., Bender, A., Weber, M., 2014. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biology 15, 545. https://doi.org/10.1186/s13059-014-0545-5

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., Kouzarides, T., 2005. Spatial Distribution of Di- and Tri-methyl Lysine 36 of Histone H3 at Active Genes. J. Biol. Chem. 280, 17732–17736. https://doi.org/10.1074/jbc.M500796200

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., Bourc'his, D., 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354, 909–912. https://doi.org/10.1126/science.aah5143

Barr, H., Hermann, A., Berger, J., Tsai, H.-H., Adie, K., Prokhortchouk, A., Hendrich, B., Bird, A., 2007. Mbd2 Contributes to DNA Methylation-Directed Repression of the Xist Gene. Molecular and Cellular Biology 27, 3750–3757. https://doi.org/10.1128/MCB.02204-06

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., Kouzarides, T., 2010. Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. Cell 143, 470–484. https://doi.org/10.1016/j.cell.2010.10.012

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., Schübeler, D., 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243–247. https://doi.org/10.1038/nature14176

Baubec, T., Ivánek, R., Lienert, F., Schübeler, D., 2013. Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. Cell 153, 480–492. https://doi.org/10.1016/j.cell.2013.03.011

Beckett, D., Kovaleva, E., Schatz, P.J., 1999. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Science 8, 921–929. https://doi.org/10.1110/ps.8.4.921

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., Lander, E.S., 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell 125, 315–326. https://doi.org/10.1016/j.cell.2006.02.041

Bird, A., 2002. DNA methylation patterns and epigenetic memory. Genes Dev. 16, 6–21. https://doi.org/10.1101/gad.947102

Bird, A., Taggart, M., Frommer, M., Miller, O.J., Macleod, D., 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40, 91–99. https://doi.org/10.1016/0092-8674(85)90312-5

Bird, A.P., 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8, 1499–1504. https://doi.org/10.1093/nar/8.7.1499

Bird, A.P., 1978. Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. Journal of Molecular Biology 118, 49–60. https://doi.org/10.1016/0022-2836(78)90243-7

Bock, I., Kudithipudi, S., Tamas, R., Kungulovski, G., Dhayalan, A., Jeltsch, A., 2011. Application of Celluspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails. BMC Biochemistry 12, 48. https://doi.org/10.1186/1471-2091-12-48

Boeke, J., Ammerpohl, O., Kegel, S., Moehren, U., Renkawitz, R., 2000. The Minimal Repression Domain of MBD2b Overlaps with the Methyl-CpG-binding Domain and Binds Directly to Sin3A*. Journal of Biological Chemistry 275, 34963–34967. https://doi.org/10.1074/jbc.M005929200

Bogdanović, O., Long, S.W., Heeringen, S.J. van, Brinkman, A.B., Gómez-Skarmeta, J.L., Stunnenberg, H.G., Jones, P.L., Veenstra, G.J.C., 2011. Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. Genome Res. 21, 1313–1327. https://doi.org/10.1101/gr.114843.110

Bonnet-Garnier, A., Feuerstein, P., Chebrout, M., Fleurot, R., Jan, H.-U., Debey, P., Beaujean, N., 2013. Genome organization and epigenetic marks in mouse germinal vesicle oocytes. Int. J. Dev. Biol. 56, 877–887. https://doi.org/10.1387/ijdb.120149ab

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., Weber, M., 2010. Targets and dynamics of promoter DNA methylation during early mouse development. Nature Genetics 42, 1093–1100. https://doi.org/10.1038/ng.708

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., Jacobsen, S.E., 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science 317, 1760–1764. https://doi.org/10.1126/science.1147939

Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B., Bestor, T.H., 2001. Dnmt3L and the Establishment of Maternal Genomic Imprints. Science 294, 2536–2539. https://doi.org/10.1126/science.1065848

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., Bell, G.W., Otte, A.P., Vidal, M., Gifford, D.K., Young, R.A., Jaenisch, R., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349–353. https://doi.org/10.1038/nature04733

Branco, M.R., King, M., Perez-Garcia, V., Bogutz, A.B., Caley, M., Fineberg, E., Lefebvre, L., Cook, S.J., Dean, W., Hemberger, M., Reik, W., 2016. Maternal DNA Methylation Regulates Early Trophoblast Development. Developmental Cell 36, 152–163. https://doi.org/10.1016/j.devcel.2015.12.027

Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., Stunnenberg, H.G., 2012. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. 22, 1128–1138. https://doi.org/10.1101/gr.133728.111

Broche, J., Kungulovski, G., Bashtrykov, P., Rathert, P., Jeltsch, A., 2020. Genome-wide investigation of the dynamic changes of epigenome modifications after global DNA methylation editing. Nucleic Acids Research. https://doi.org/10.1093/nar/gkaa1169

Cantone, I., Fisher, A.G., 2013. Epigenetic programming and reprogramming during development. Nat Struct Mol Biol 20, 282–289. https://doi.org/10.1038/nsmb.2489

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., Zhang, Y., 2002. Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science 298, 1039–1043. https://doi.org/10.1126/science.1076997

Cardoso, M.C., Leonhardt, H., 1999. DNA Methyltransferase Is Actively Retained in the Cytoplasm during Early Development. J Cell Biol 147, 25–32.

Chédin, F., Lieber, M.R., Hsieh, C.-L., 2002. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. PNAS 99, 16916–16921. https://doi.org/10.1073/pnas.262443999

Chen, R., Wu, X., Jiang, L., Zhang, Y., 2017. Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. Cell Reports 18, 3227–3241. https://doi.org/10.1016/j.celrep.2017.03.004

Chen, T., Tsujimoto, N., Li, E., 2004. The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin. Mol. Cell. Biol. 24, 9048–9058. https://doi.org/10.1128/MCB.24.20.9048-9058.2004

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., Li, E., 2003. Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. Mol. Cell. Biol. 23, 5594–5605. https://doi.org/10.1128/MCB.23.16.5594-5605.2003

Chen, T., Ueda, Y., Xie, S., Li, E., 2002. A Novel Dnmt3a Isoform Produced from an Alternative Promoter Localizes to Euchromatin and Its Expression Correlates with Activede Novo Methylation. J. Biol. Chem. 277, 38746–38754. https://doi.org/10.1074/jbc.M205312200

Chen, Z.-X., Mann, J.R., Hsieh, C.-L., Riggs, A.D., Chédin, F., 2005. Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J. Cell. Biochem. 95, 902–917. https://doi.org/10.1002/jcb.20447

Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., Kelsey, G., 2009. Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev. 23, 105–117. https://doi.org/10.1101/gad.495809

Christian, D.L., Wu, D.Y., Martin, J.R., Moore, J.R., Liu, Y.R., Clemens, A.W., Nettles, S.A., Kirkland, N.M., Papouin, T., Hill, C.A., Wozniak, D.F., Dougherty, J.D., Gabel, H.W., 2020. DNMT3A Haploinsufficiency Results in Behavioral Deficits and Global Epigenomic Dysregulation Shared across Neurodevelopmental Disorders. Cell Reports 33, 108416. https://doi.org/10.1016/j.celrep.2020.108416

Chuang, L.S.-H., Ian, H.-I., Koh, T.-W., Ng, H.-H., Xu, G., Li, B.F.L., 1997. Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1. Science 277, 1996–2000. https://doi.org/10.1126/science.277.5334.1996

Clark, S.J., Smallwood, S.A., Lee, H.J., Krueger, F., Reik, W., Kelsey, G., 2017. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). Nat Protoc 12, 534–547. https://doi.org/10.1038/nprot.2016.187

Clements, E.G., Mohammad, H.P., Leadem, B.R., Easwaran, H., Cai, Y., Van Neste, L., Baylin, S.B., 2012. DNMT1 modulates gene expression without its catalytic activity partially through its interactions with histone-modifying enzymes. Nucleic Acids Res 40, 4334–4346. https://doi.org/10.1093/nar/gks031

Clemson, C.M., McNeil, J.A., Willard, H.F., Lawrence, J.B., 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. Journal of Cell Biology 132, 259–275. https://doi.org/10.1083/jcb.132.3.259

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L.K., Bartolomei, M.S., Rambow, F., Bassi, M.R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A.J., Matsumoto, Y., Kobi, D., Davidson, I., Alberti, C., Larue, L., Bellacosa, A., 2011. Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. Cell 146, 67–79. https://doi.org/10.1016/j.cell.2011.06.020

Dahlet, T., Lleida, A.A., Adhami, H.A., Dumas, M., Bender, A., Ngondo, R.P., Tanguy, M., Vallet, J., Auclair, G., Bardet, A.F., Weber, M., 2020. Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. Nat Commun 11, 1–14. https://doi.org/10.1038/s41467-020-16919-w

Datta, J., Majumder, S., Bai, S., Ghoshal, K., Kutay, H., Smith, D.S., Crabb, J.W., Jacob, S.T., 2005. Physical and functional interaction of DNA methyltransferase 3A with Mbd3 and Brg1 in mouse lymphosarcoma cells. Cancer Res 65, 10891–10900. https://doi.org/10.1158/0008-5472.CAN-05-1455

Deplus, R., Brenner, C., Burgers, W.A., Putmans, P., Kouzarides, T., Launoit, Y. de, Fuks, F., 2002. Dnmt3L is a transcriptional repressor that recruits histone deacetylase. Nucleic Acids Research 30, 3831–3838. https://doi.org/10.1093/nar/gkf509

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., Jeltsch, A., 2010. The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. J. Biol. Chem. 285, 26114–26120. https://doi.org/10.1074/jbc.M109.089433

Du, Q., Luu, P.-L., Stirzaker, C., Clark, S.J., 2015. Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics 7, 1051–1073. https://doi.org/10.2217/epi.15.39

Du, W., Dong, Q., Zhang, Z., Liu, B., Zhou, T., Xu, R., Wang, H., Zhu, B., Li, Y., 2019. Stella protein facilitates DNA demethylation by disrupting the chromatin association of the RING finger–type E3 ubiquitin ligase UHRF1. J. Biol. Chem. 294, 8907–8917. https://doi.org/10.1074/jbc.RA119.008008

Dukatz, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P., Jeltsch, A., 2019. H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction. Journal of Molecular Biology 431, 5063–5074. https://doi.org/10.1016/j.jmb.2019.09.006

Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., McCune, R.A., Gehrke, C., 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res 10, 2709–2721. https://doi.org/10.1093/nar/10.8.2709

Ehrlich, M., Norris, K.F., Wang, R.Y., Kuo, K.C., Gehrke, C.W., 1986. DNA cytosine methylation and heat-induced deamination. Biosci Rep 6, 387–393. https://doi.org/10.1007/BF01116426

Ehrlich, M., Wang, R.Y., 1981. 5-Methylcytosine in eukaryotic DNA. Science 212, 1350–1357. https://doi.org/10.1126/science.6262918

Ehrlich, M., Zhang, X.-Y., Inamdar, N.M., 1990. Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. Mutation Research/Reviews in Genetic Toxicology 238, 277–286. https://doi.org/10.1016/0165-1110(90)90019-8

Emperle, M., Dukatz, M., Kunert, S., Holzer, K., Rajavelu, A., Jurkowska, R.Z., Jeltsch, A., 2018. The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A/R882H complexes. Sci Rep 8, 1–9. https://doi.org/10.1038/s41598-018-31635-8

Emperle, M., Rajavelu, A., Reinhardt, R., Jurkowska, R.Z., Jeltsch, A., 2014. Cooperative DNA Binding and Protein/DNA Fiber Formation Increases the Activity of the Dnmt3a

DNA Methyltransferase. J. Biol. Chem. 289, 29602–29613. https://doi.org/10.1074/jbc.M114.572032

Fairhead, M., Howarth, M., 2015. Site-specific biotinylation of purified proteins using BirA. Methods Mol Biol 1266, 171–184. https://doi.org/10.1007/978-1-4939-2272-7_12

Feng, J., Chang, H., Li, E., Fan, G., 2005. Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. Journal of Neuroscience Research 79, 734–746. https://doi.org/10.1002/jnr.20404

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., Ukomadu, C., Sadler, K.C., Pradhan, S., Pellegrini, M., Jacobsen, S.E., 2010. Conservation and divergence of methylation patterning in plants and animals. PNAS 107, 8689–8694. https://doi.org/10.1073/pnas.1002720107

Fuks, F., Burgers, W.A., Godin, N., Kasai, M., Kouzarides, T., 2001. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. The EMBO Journal 20, 2536–2544. https://doi.org/10.1093/emboj/20.10.2536

Gahurova, L., Tomizawa, S., Smallwood, S.A., Stewart-Morgan, K.R., Saadeh, H., Kim, J., Andrews, S.R., Chen, T., Kelsey, G., 2017. Transcription and chromatin determinants of de novo DNA methylation timing in oocytes. Epigenetics & Chromatin 10, 25. https://doi.org/10.1186/s13072-017-0133-5

Galupa, R., Heard, E., 2015. X-chromosome inactivation: new insights into cis and trans regulation. Current Opinion in Genetics & Development, Genome architecture and expression 31, 57–66. https://doi.org/10.1016/j.gde.2015.04.002

Gama-Sosa, M.A., Midgett, R.M., Slagel, V.A., Githens, S., Kuo, K.C., Gehrke, C.W., Ehrlich, M., 1983. Tissue-specific differences in DNA methylation in various mammals. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 740, 212–219. https://doi.org/10.1016/0167-4781(83)90079-9

Gao, F., Ji, G., Gao, Z., Han, X., Ye, M., Yuan, Z., Luo, H., Huang, X., Natarajan, K., Wang, J., Yang, H., Zhang, X., 2014. Direct ChIP-bisulfite sequencing reveals a role of H3K27me3 mediating aberrant hypermethylation of promoter CpG islands in cancer cells. Genomics 103, 204–210. https://doi.org/10.1016/j.ygeno.2013.12.006

Gao, L., Emperle, M., Guo, Y., Grimm, S.A., Ren, W., Adam, S., Uryu, H., Zhang, Z.-M., Chen, D., Yin, J., Dukatz, M., Anteneh, H., Jurkowska, R.Z., Lu, J., Wang, Y., Bashtrykov, P., Wade, P.A., Wang, G.G., Jeltsch, A., Song, J., 2020. Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat Commun 11, 1–14. https://doi.org/10.1038/s41467-020-17109-4

Ge, Y.-Z., Pu, M.-T., Gowher, H., Wu, H.-P., Ding, J.-P., Jeltsch, A., Xu, G.-L., 2004. Chromatin Targeting of de Novo DNA Methyltransferases by the PWWP Domain. J. Biol. Chem. 279, 25447–25454. https://doi.org/10.1074/jbc.M312296200

Geiman, T.M., Sankpal, U.T., Robertson, A.K., Zhao, Yingxin, Zhao, Yingming, Robertson, K.D., 2004. DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. Biochem Biophys Res Commun 318, 544–555. https://doi.org/10.1016/j.bbrc.2004.04.058

Gelfman, S., Cohen, N., Yearim, A., Ast, G., 2013. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon–intron structure. Genome Res. 23, 789–799. https://doi.org/10.1101/gr.143503.112

Gierlinski, M., Gastaldello, F., Cole, C., Barton, G.J., 2018. Proteus: an R package for downstream analysis of MaxQuant output. bioRxiv 416511. https://doi.org/10.1101/416511

Gong, T., Gu, X., Liu, Y.-T., Zhou, Z., Zhang, L.-L., Wen, Y., Zhong, W.-L., Xu, G.-L., Zhou, J.-Q., 2020. Both combinatorial K4me0-K36me3 marks on sister histone H3s of a nucleosome are required for Dnmt3a-Dnmt3L mediated de novo DNA methylation. Journal of Genetics and Genomics 47, 105–114. https://doi.org/10.1016/j.jgg.2019.12.006

Gonzalo, S., Jaco, I., Fraga, M.F., Chen, T., Li, E., Esteller, M., Blasco, M.A., 2006. DNA methyltransferases control telomere length and telomere recombination in mammalian cells. Nat Cell Biol 8, 416–424. https://doi.org/10.1038/ncb1386

Gopalakrishnan, S., Sullivan, B.A., Trazzi, S., Della Valle, G., Robertson, K.D., 2009. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum Mol Genet 18, 3178–3193. https://doi.org/10.1093/hmg/ddp256

Gowher, H., Liebert, K., Hermann, A., Xu, G., Jeltsch, A., 2005. Mechanism of Stimulation of Catalytic Activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J. Biol. Chem. 280, 13341–13348. https://doi.org/10.1074/jbc.M413412200

Greenberg, M.V.C., Bourc'his, D., 2019. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 20, 590–607. https://doi.org/10.1038/s41580-019-0159-6

Greenberg, M.V.C., Glaser, J., Borsos, M., Marjou, F.E., Walter, M., Teissandier, A., Bourc'his, D., 2017. Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. Nat Genet 49, 110–118. https://doi.org/10.1038/ng.3718

Gu, H., Zou, Y.-R., Rajewsky, K., 1993. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Cell 73, 1155–1164. https://doi.org/10.1016/0092-8674(93)90644-6

Gu, T., Lin, X., Cullen, S.M., Luo, M., Jeong, M., Estecio, M., Shen, J., Hardikar, S., Sun, D., Su, J., Rux, D., Guzman, A., Lee, M., Qi, L.S., Chen, J.-J., Kyba, M., Huang, Y., Chen, T., Li, W., Goodell, M.A., 2018. DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. Genome Biol. 19, 88. https://doi.org/10.1186/s13059-018-1464-7

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., Iqbal, K., Shi, Y.G., Deng, Z., Szabó, P.E., Pfeifer, G.P., Li, J., Xu, G.-L., 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477, 606–610. https://doi.org/10.1038/nature10443

Guibert, S., Forné, T., Weber, M., 2012. Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res 22, 633–641. https://doi.org/10.1101/gr.130997.111

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., Walsh, C.P., Li, J., Tang, F., Xu, G.-L., 2014. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. Cell Stem Cell 15, 447–459. https://doi.org/10.1016/j.stem.2014.08.003

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G., Song, H., 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci 17, 215–222. https://doi.org/10.1038/nn.3607

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., Tian, C., Wang, J., Cong, Y., Xu, Y., 2015. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640–644. https://doi.org/10.1038/nature13899

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., Surani, M.A., 2010. Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway. Science 329, 78–82. https://doi.org/10.1126/science.1187945

Han, M., Li, Jialun, Cao, Y., Huang, Y., Li, W., Zhu, H., Zhao, Q., Han, J.-D.J., Wu, Q., Li, Jiwen, Feng, J., Wong, J., 2020. A role for LSH in facilitating DNA methylation by DNMT1 through enhancing UHRF1 chromatin association. Nucleic Acids Research 48, 12116–12134. https://doi.org/10.1093/nar/gkaa1003

Hanna, C.W., Taudt, A., Huang, J., Gahurova, L., Kranz, A., Andrews, S., Dean, W., Stewart, A.F., Colomé-Tatché, M., Kelsey, G., 2018. MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. Nature Structural & Molecular Biology 25, 73. https://doi.org/10.1038/s41594-017-0013-5

Hata, K., Okano, M., Lei, H., Li, E., 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development 129, 1983–1993.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., Xu, G.-L., 2011. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science 333, 1303–1307. https://doi.org/10.1126/science.1210944

Hendrich, B., Bird, A., 1998. Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. Molecular and Cellular Biology 18, 6538–6547. https://doi.org/10.1128/MCB.18.11.6538

Henikoff, S., Smith, M.M., 2015. Histone Variants and Epigenetics. Cold Spring Harb Perspect Biol 7. https://doi.org/10.1101/cshperspect.a019364

Hermann, A., Goyal, R., Jeltsch, A., 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. J. Biol. Chem. 279, 48350–48359. https://doi.org/10.1074/jbc.M403427200

Heyn, P., Logan, C.V., Fluteau, A., Challis, R.C., Auchynnikava, T., Martin, C.-A., Marsh, J.A., Taglini, F., Kilanowski, F., Parry, D.A., Cormier-Daire, V., Fong, C.-T., Gibson, K., Hwa, V., Ibáñez, L., Robertson, S.P., Sebastiani, G., Rappsilber, J., Allshire, R.C., Reijns, M.A.M., Dauber, A., Sproul, D., Jackson, A.P., 2019. Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. Nature Genetics 51, 96. https://doi.org/10.1038/s41588-018-0274-x

Holliday, R., Grigg, G.W., 1993. DNA methylation and mutation. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, Special Issue In Memory of Max Clark, a Pioneer in Fundamental Mutation Reserach 285, 61–67. https://doi.org/10.1016/0027-5107(93)90052-H

Hsieh, C.L., 1994. Dependence of transcriptional repression on CpG methylation density. Molecular and Cellular Biology 14, 5487–5494. https://doi.org/10.1128/MCB.14.8.5487

Huang, Y.-H., Su, J., Lei, Y., Brunetti, L., Gundry, M.C., Zhang, X., Jeong, M., Li, W., Goodell, M.A., 2017. DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. Genome Biol 18. https://doi.org/10.1186/s13059-017-1306-z

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., Bird, A.P., 2010. Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLoS Genet 6, e1001134. https://doi.org/10.1371/journal.pgen.1001134

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., Zhang, Y., 2011. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science 333, 1300–1303. https://doi.org/10.1126/science.1210597

Jain, D., Meydan, C., Lange, J., Bouuaert, C.C., Lailler, N., Mason, C.E., Anderson, K.V., Keeney, S., 2017. rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. PLOS Genetics 13, e1006964. https://doi.org/10.1371/journal.pgen.1006964

Jeltsch, A., 2002. Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. ChemBioChem 3, 274–293. https://doi.org/10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S

Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G.A., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R., Kim, S.-B., Yang, L., Ko, M., Chen, R., Göttgens, B., Lee, J.-S., Gunaratne, P., Godley, L.A., Darlington, G.J., Rao, A., Li, W., Goodell, M.A., 2014. Large conserved domains of low DNA methylation maintained by Dnmt3a. Nature Genetics 46, 17–23. https://doi.org/10.1038/ng.2836

Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S., Jones, P.A., 2009. Selective Anchoring of DNA Methyltransferases 3A and 3B to Nucleosomes Containing Methylated DNA. Molecular and Cellular Biology 29, 5366–5376. https://doi.org/10.1128/MCB.00484-09

Jermann, P., Hoerner, L., Burger, L., Schübeler, D., 2014. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. PNAS 111, E3415–E3421. https://doi.org/10.1073/pnas.1400672111

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., Cheng, X., 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449, 248–251. https://doi.org/10.1038/nature06146

Jones, P.L., Veenstra, G.J.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., Wolffe, A.P., 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet 19, 187–191. https://doi.org/10.1038/561

Jurkowska, R.Z., Siddique, A.N., Jurkowski, T.P., Jeltsch, A., 2011. Approaches to Enzyme and Substrate Design of the Murine Dnmt3a DNA Methyltransferase. ChemBioChem 12, 1589–1594. https://doi.org/10.1002/cbic.201000673

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., Saitou, M., 2013. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J 32, 340–353. https://doi.org/10.1038/emboj.2012.331

Kaneda, M., Hirasawa, R., Chiba, H., Okano, M., Li, E., Sasaki, H., 2010. Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. Genes to Cells 15, 169–179. https://doi.org/10.1111/j.1365-2443.2009.01374.x

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., Sasaki, H., 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429, 900–903. https://doi.org/10.1038/nature02633

Kareta, M.S., Botello, Z.M., Ennis, J.J., Chou, C., Chédin, F., 2006. Reconstitution and Mechanism of the Stimulation of de Novo Methylation by Human DNMT3L. J. Biol. Chem. 281, 25893–25902. https://doi.org/10.1074/jbc.M603140200

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., Suzuki, Y., Kono, T., 2012. Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. PLOS Genetics 8, e1002440. https://doi.org/10.1371/journal.pgen.1002440

Kungulovski, G., Kycia, I., Tamas, R., Jurkowska, R.Z., Kudithipudi, S., Henry, C., Reinhardt, R., Labhart, P., Jeltsch, A., 2014. Application of histone modification-specific interaction domains as an alternative to antibodies. Genome Res. 24, 1842–1853. https://doi.org/10.1101/gr.170985.113

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H., Nakano, T., 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 22, 908–917. https://doi.org/10.1101/gad.1640708

Lee, H.J., Hore, T.A., Reik, W., 2014. Reprogramming the Methylome: Erasing Memory and Creating Diversity. Cell Stem Cell 14, 710–719. https://doi.org/10.1016/j.stem.2014.05.008

Lee, J.-H., Park, S.-J., Nakai, K., 2017. Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. Sci Rep 7. https://doi.org/10.1038/s41598-017-11800-1

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., Peters, A.H.F.M., 2003. Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. Current Biology 13, 1192–1200. https://doi.org/10.1016/S0960-9822(03)00432-9

Li, B.-Z., Huang, Z., Cui, Q.-Y., Song, X.-H., Du, L., Jeltsch, A., Chen, P., Li, G., Li, E., Xu, G.-L., 2011. Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell Res 21, 1172–1181. https://doi.org/10.1038/cr.2011.92

Li, E., Bestor, T.H., Jaenisch, R., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69, 915–926. https://doi.org/10.1016/0092-8674(92)90611-F

Li, G., Zhang, W., Baker, M.S., Laritsky, E., Mattan-Hung, N., Yu, D., Kunde-Ramamoorthy, G., Simerly, R.B., Chen, R., Shen, L., Waterland, R.A., 2014. Major epigenetic development distinguishing neuronal and non-neuronal cells occurs postnatally in the murine hypothalamus. Hum Mol Genet 23, 1579–1590. https://doi.org/10.1093/hmg/ddt548

Li, J.-Y., Pu, M.-T., Hirasawa, R., Li, B.-Z., Huang, Y.-N., Zeng, R., Jing, N.-H., Chen, T., Li, E., Sasaki, H., Xu, G.-L., 2007a. Synergistic Function of DNA Methyltransferases Dnmt3a and Dnmt3b in the Methylation of Oct4 and Nanog. Molecular and Cellular Biology 27, 8748–8759. https://doi.org/10.1128/MCB.01380-07

Li, J.-Y., Pu, M.-T., Hirasawa, R., Li, B.-Z., Huang, Y.-N., Zeng, R., Jing, N.-H., Chen, T., Li, E., Sasaki, H., Xu, G.-L., 2007b. Synergistic Function of DNA Methyltransferases Dnmt3a and Dnmt3b in the Methylation of Oct4 and Nanog. Molecular and Cellular Biology 27, 8748–8759. https://doi.org/10.1128/MCB.01380-07

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., Ferguson-Smith, A.C., 2008. A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. Developmental Cell 15, 547–557. https://doi.org/10.1016/j.devcel.2008.08.014

Li, Y., Li, J., 2019. Technical advances contribute to the study of genomic imprinting. PLOS Genetics 15, e1008151. https://doi.org/10.1371/journal.pgen.1008151

Li, Y., Zheng, H., Wang, Q., Zhou, C., Wei, L., Liu, X., Zhang, W., Zhang, Y., Du, Z., Wang, X., Xie, W., 2018. Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. Genome Biology 19, 18. https://doi.org/10.1186/s13059-018-1390-8

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., Schübeler, D., 2011. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet 43, 1091–1097. https://doi.org/10.1038/ng.946

Lindroth, A.M., Park, Y.J., McLean, C.M., Dokshin, G.A., Persson, J.M., Herman, H., Pasini, D., Miró, X., Donohoe, M.E., Lee, J.T., Helin, K., Soloway, P.D., 2008. Antagonism between DNA and H3K27 Methylation at the Imprinted Rasgrf1 Locus. PLOS Genetics 4, e1000145. https://doi.org/10.1371/journal.pgen.1000145

Lio, C.-W.J., Rao, A., 2019. TET Enzymes and 5hmC in Adaptive and Innate Immune Systems. Front. Immunol. 10. https://doi.org/10.3389/fimmu.2019.00210

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J.C., Rao, A., Esteller, M., He, C., Haghighi, F.G., Sejnowski, T.J., Behrens, M.M., Ecker, J.R., 2013. Global Epigenomic Reconfiguration During Mammalian Brain Development. Science 341, 1237905. https://doi.org/10.1126/science.1237905

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., Ecker, J.R., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315–322. https://doi.org/10.1038/nature08514

Lucifero, D., La Salle, S., Bourc'his, D., Martel, J., Bestor, T.H., Trasler, J.M., 2007. Coordinate regulation of DNA methyltransferase expression during oogenesis. BMC Dev Biol 7, 36. https://doi.org/10.1186/1471-213X-7-36

Mallona, I., Ilie, I.M., Karemaker, I.D., Butz, S., Manzo, M., Caflisch, A., Baubec, T., 2020. Flanking sequence preference modulates de novo DNA methylation in the mouse genome. Nucleic Acids Research. https://doi.org/10.1093/nar/gkaa1168

Manzo, M., Wirz, J., Ambrosi, C., Villaseñor, R., Roschitzki, B., Baubec, T., 2017. Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. EMBO J. 36, 3421–3434. https://doi.org/10.15252/embj.201797038

Mao, S.-Q., Cuesta, S.M., Tannahill, D., Balasubramanian, S., 2020. Genome-wide DNA Methylation Signatures Are Determined by DNMT3A/B Sequence Preferences. Biochemistry 59, 2541–2550. https://doi.org/10.1021/acs.biochem.0c00339

Marshall, J.J., Mason, J.O., 2019. Mouse vs man: Organoid models of brain development & disease. Brain Res 1724, 146427. https://doi.org/10.1016/j.brainres.2019.146427

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V.M., Rowitch, D.H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S.J.M., Haussler, D., Marra, M.A., Hirst, M., Wang, T., Costello, J.F., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466, 253–257. https://doi.org/10.1038/nature09165

Mauser, R., Kungulovski, G., Keup, C., Reinhardt, R., Jeltsch, A., 2017. Application of dual reading domains as novel reagents in chromatin biology reveals a new H3K9me3 and H3K36me2/3 bivalent chromatin state. Epigenetics & Chromatin 10, 45. https://doi.org/10.1186/s13072-017-0153-1

Mayer, W., Niveleau, A., Walter, J., Fundele, R., Haaf, T., 2000. Demethylation of the zygotic paternal genome. Nature 403, 501–502. https://doi.org/10.1038/35000656

Mellid, S., Coloma, J., Calsina, B., Monteagudo, M., Roldán-Romero, J.M., Santos, M., Leandro-García, L.J., Lanillos, J., Martínez-Montes, Á.M., Rodríguez-Antona, C., Montero-Conde, C., Martínez-López, J., Ayala, R., Matias-Guiu, X., Robledo, M., Cascón, A., 2020. Novel DNMT3A Germline Variant in a Patient with Multiple Paragangliomas and Papillary Thyroid Carcinoma. Cancers (Basel) 12. https://doi.org/10.3390/cancers12113304

Mendoza, A. de, Poppe, D., Buckberry, S., Pflueger, J., Albertin, C.B., Daish, T., Bertrand, S., Calle-Mustienes, E. de la, Gómez-Skarmeta, J.L., Nery, J.R., Ecker, J.R., Baer, B., Ragsdale, C.W., Grützner, F., Escriva, H., Venkatesh, B., Bogdanovic, O., Lister, R., 2021. The emergence of the brain non-CpG methylation system in vertebrates. Nat Ecol Evol 1–10. https://doi.org/10.1038/s41559-020-01371-2

Mihalache, L., Gherasim, A., Niță, O., Ungureanu, M.C., Pădureanu, S.S., Gavril, R.S., Arhire, L.I., 2016. Effects of ghrelin in energy balance and body weight homeostasis. HORMONES 15, 186–196. https://doi.org/10.14310/horm.2002.1672

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., Schübeler, D., 2008. Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. Molecular Cell 30, 755–766. https://doi.org/10.1016/j.molcel.2008.05.007

Morselli, M., Pastor, W.A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A.T., Ottonello, S., Jacobsen, S.E., Pellegrini, M., 2015. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Elife 4, e06205. https://doi.org/10.7554/eLife.06205

Mulholland, C.B., Nishiyama, A., Ryan, J., Nakamura, R., Yiğit, M., Glück, I.M., Trummer, C., Qin, W., Bartoschek, M.D., Traube, F.R., Parsa, E., Ugur, E., Modic, M., Acharya, A., Stolz, P., Ziegenhain, C., Wierer, M., Enard, W., Carell, T., Lamb, D.C., Takeda, H., Nakanishi, M., Bultmann, S., Leonhardt, H., 2020. Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. Nat Commun 11, 1–24. https://doi.org/10.1038/s41467-020-19603-1

Murphy, P.J., Cipriany, B.R., Wallin, C.B., Ju, C.Y., Szeto, K., Hagarman, J.A., Benitez, J.J., Craighead, H.G., Soloway, P.D., 2013. Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells. PNAS 110, 7772–7777. https://doi.org/10.1073/pnas.1218495110

Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., Heras, J.I. de L., Stancheva, I., 2011. LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Res. 21, 83–94. https://doi.org/10.1101/gr.108498.110

Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393, 386–389. https://doi.org/10.1038/30764

Neddermann, P., Jiricny, J., 1993. The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. J Biol Chem 268, 21218–21224.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., Oliviero, S., 2017. Intragenic DNA methylation prevents spurious transcription initiation. Nature 543, 72–77. https://doi.org/10.1038/nature21373

Ng, H.-H., Jeppesen, P., Bird, A., 2000. Active Repression of Methylated Genes by the Chromosomal Protein MBD1. Molecular and Cellular Biology 20, 1394–1406. https://doi.org/10.1128/MCB.20.4.1394-1406.2000

Noh, K.-M., Wang, H., Kim, H.R., Wenderski, W., Fang, F., Li, C.H., Dewell, S., Hughes, S.H., Melnick, A.M., Patel, D.J., Li, H., Allis, C.D., 2015. Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. Molecular Cell 59, 89–103. https://doi.org/10.1016/j.molcel.2015.05.017

Norris, D.P., Patel, D., Kay, G.F., Penny, G.D., Brockdorff, N., Sheardown, S.A., Rastan, S., 1994. Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell 77, 41–51. https://doi.org/10.1016/0092-8674(94)90233-x

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., Berman, D.M., Jenuwein, T., Pruitt, K., Sharkis, S.J., Watkins, D.N., Herman, J.G., Baylin, S.B., 2007. A stem cell–like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 39, 237–242. https://doi.org/10.1038/ng1972

Ohno, R., Nakayama, M., Naruse, C., Okashita, N., Takano, O., Tachibana, M., Asano, M., Saitou, M., Seki, Y., 2013. A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. Development 140, 2892–2903. https://doi.org/10.1242/dev.093229

Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell 99, 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., Cheng, X., Bestor, T.H., 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714–717. https://doi.org/10.1038/nature05987

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., Walter, J., 2000. Active demethylation of the paternal genome in the mouse zygote. Current Biology 10, 475–478. https://doi.org/10.1016/S0960-9822(00)00448-6

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., Shirakawa, M., 2009. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. EMBO reports 10, 1235–1241. https://doi.org/10.1038/embor.2009.218

Proudhon, C., Duffié, R., Ajjan, S., Cowley, M., Iranzo, J., Carbajosa, G., Saadeh, H., Holland, M.L., Oakey, R.J., Rakyan, V.K., Schulz, R., Bourc'his, D., 2012. Protection against De Novo Methylation Is Instrumental in Maintaining Parent-of-Origin Methylation Inherited from the Gametes. Molecular Cell 47, 909–920. https://doi.org/10.1016/j.molcel.2012.07.010

Qin, S., Min, J., 2014. Structure and function of the nucleosome-binding PWWP domain. Trends in Biochemical Sciences 39, 536–547. https://doi.org/10.1016/j.tibs.2014.09.001

Qiu, C., Sawada, K., Zhang, X., Cheng, X., 2002. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Mol Biol 9, 217–224. https://doi.org/10.1038/nsb759

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., Trono, D., 2011. In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. Molecular Cell 44, 361–372. https://doi.org/10.1016/j.molcel.2011.08.032

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., Jaenisch, R., 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97, 5237–5242.

Ranke, M.B., Wit, J.M., 2018. Growth hormone — past, present and future. Nat Rev Endocrinol 14, 285–300. https://doi.org/10.1038/nrendo.2018.22

Remacha, L., Currás-Freixes, M., Torres-Ruiz, R., Schiavi, F., Torres-Pérez, R., Calsina, B., Letón, R., Comino-Méndez, I., Roldán-Romero, J.M., Montero-Conde, C., Santos, M., Pérez, L.I., Pita, G., Alonso, M.R., Honrado, E., Pedrinaci, S., Crespo-Facorro, B., Percesepe, A., Falcioni, M., Rodríguez-Perales, S., Korpershoek, E., Ramón-Maiques, S., Opocher, G., Rodríguez-Antona, C., Robledo, M., Cascón, A., 2018. Gain-of-function mutations in DNMT3A in patients with paraganglioma. Genetics in Medicine 20, 1644–1651. https://doi.org/10.1038/s41436-018-0003-y

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., Klug, A., 1984. Structure of the nucleosome core particle at 7 Å resolution. Nature 311, 532–537. https://doi.org/10.1038/311532a0

Rondelet, G., Dal Maso, T., Willems, L., Wouters, J., 2016. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. Journal of Structural Biology 194, 357–367. https://doi.org/10.1016/j.jsb.2016.03.013

Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Pàldi, A., Viegas-Péquignot, E., 1998. Chromosome methylation patterns during mammalian preimplantation development. Genes Dev 12, 2108–2113.

Rountree, M.R., Bachman, K.E., Baylin, S.B., 2000. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet 25, 269–277. https://doi.org/10.1038/77023

Roy, P.H., Weissbach, A., 1975. DNA methylase from HeLa cell nuclei. Nucleic Acids Res 2, 1669–1684. https://doi.org/10.1093/nar/2.10.1669

Rutledge, C.E., Thakur, A., O'Neill, K.M., Irwin, R.E., Sato, S., Hata, K., Walsh, C.P., 2014. Ontogeny, conservation and functional significance of maternally inherited DNA methylation at two classes of non-imprinted genes. Development 141, 1313–1323. https://doi.org/10.1242/dev.104646

Sanchez-Delgado, M., Court, F., Vidal, E., Medrano, J., Monteagudo-Sánchez, A., Martin-Trujillo, A., Tayama, C., Iglesias-Platas, I., Kondova, I., Bontrop, R., Poo-Llanillo, M.E., Marques-Bonet, T., Nakabayashi, K., Simón, C., Monk, D., 2016. Human Oocyte-Derived Methylation Differences Persist in the Placenta Revealing Widespread Transient Imprinting. PLOS Genetics 12, e1006427. https://doi.org/10.1371/journal.pgen.1006427

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., Kouzarides, T., 2002. Active genes are tri-methylated at K4 of histone H3. Nature 419, 407–411. https://doi.org/10.1038/nature01080

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., Bergman, Y., Simon, I., Cedar, H., 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 39, 232–236. https://doi.org/10.1038/ng1950

Schübeler, D., 2015. Function and information content of DNA methylation. Nature 517, 321–326. https://doi.org/10.1038/nature14192

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., Lin, S., Lin, Y., Jung, I., Schmitt, A.D., Selvaraj, S., Ren, B., Sejnowski, T.J., Wang, W., Ecker, J.R., 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523, 212–216. https://doi.org/10.1038/nature14465

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., Reik, W., 2012. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. Molecular Cell 48, 849–862. https://doi.org/10.1016/j.molcel.2012.11.001

Sendžikaitė, G., Hanna, C.W., Stewart-Morgan, K.R., Ivanova, E., Kelsey, G., 2019. A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. Nature Communications 10, 1884. https://doi.org/10.1038/s41467-019-09713-w

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., Koseki, H., 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450, 908–912. https://doi.org/10.1038/nature06397

Shayevitch, R., Askayo, D., Keydar, I., Ast, G., 2018. The importance of DNA methylation of exons on alternative splicing. RNA 24, 1351–1362. https://doi.org/10.1261/rna.064865.117

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., Zhang, Y., 2014. Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. Cell Stem Cell 15, 459–471. https://doi.org/10.1016/j.stem.2014.09.002

Shirane, K., Miura, F., Ito, T., Lorincz, M.C., 2020. NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. Nat Genet 1–11. https://doi.org/10.1038/s41588-020-0689-z

Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., Sasaki, H., 2013. Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. PLOS Genetics 9, e1003439. https://doi.org/10.1371/journal.pgen.1003439

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., Oberdoerffer, S., 2011. CTCF-promoted RNA

polymerase II pausing links DNA methylation to splicing. Nature 479, 74–79. https://doi.org/10.1038/nature10442

Siegfried, Z., Eden, S., Mendelsohn, M., Feng, X., Tsuberi, B.-Z., Cedar, H., 1999. DNA methylation represses transcription in vivo. Nat Genet 22, 203–206. https://doi.org/10.1038/9727

Silva, J.C., Gorenstein, M.V., Li, G.-Z., Vissers, J.P.C., Geromanos, S.J., 2006. Absolute Quantification of Proteins by LCMS ^{E}: A Virtue of Parallel ms Acquisition. Mol Cell Proteomics 5, 144–156. https://doi.org/10.1074/mcp.M500230-MCP200

Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., Kelsey, G., 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43, 811–814. https://doi.org/10.1038/ng.864

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., Meissner, A., 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature 484, 339–344. https://doi.org/10.1038/nature10960

Smith, Z.D., Shi, J., Gu, H., Donaghey, J., Clement, K., Cacchiarelli, D., Gnirke, A., Michor, F., Meissner, A., 2017. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. Nature 549, 543–547. https://doi.org/10.1038/nature23891

Song, J., Rechkoblit, O., Bestor, T.H., Patel, D.J., 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science 331, 1036–1040. https://doi.org/10.1126/science.1195380

Song, J., Teplova, M., Ishibe-Murakami, S., Patel, D.J., 2012. Structure-Based Mechanistic Insights into DNMT1-Mediated Maintenance DNA Methylation. Science 335, 709–712. https://doi.org/10.1126/science.1214453

Statham, A.L., Robinson, M.D., Song, J.Z., Coolen, M.W., Stirzaker, C., Clark, S.J., 2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res 22, 1120–1127. https://doi.org/10.1101/gr.132076.111

Stewart, K.R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S., Smallwood, S.A., Chen, T., Kelsey, G., 2015. Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. Genes Dev. 29, 2449–2462. https://doi.org/10.1101/gad.271353.115

Stirzaker, C., Song, J.Z., Ng, W., Du, Q., Armstrong, N.J., Locke, W.J., Statham, A.L., French, H., Pidsley, R., Valdes-Mora, F., Zotenko, E., Clark, S.J., 2017. Methyl-CpG-binding protein MBD2 plays a key role in maintenance and spread of DNA methylation at CpG islands and shores in cancer. Oncogene 36, 1328–1338. https://doi.org/10.1038/onc.2016.297

Su, J., Huang, Y.-H., Cui, X., Wang, X., Zhang, X., Lei, Y., Xu, J., Lin, X., Chen, K., Lv, J., Goodell, M.A., Li, W., 2018. Homeobox oncogene activation by pan-cancer DNA hypermethylation. Genome Biology 19, 108. https://doi.org/10.1186/s13059-018-1492-3

Suetake, I., Hayata, D., Tajima, S., 2006. The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. J Biochem 140, 763–776. https://doi.org/10.1093/jb/mvj210

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., Tajima, S., 2004. DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. J. Biol. Chem. 279, 27816–27823. https://doi.org/10.1074/jbc.M400181200

Sun, Y., Butte, N.F., Garcia, J.M., Smith, R.G., 2008. Characterization of Adult Ghrelin and Ghrelin Receptor Knockout Mice under Positive and Negative Energy Balance. Endocrinology 149, 843–850. https://doi.org/10.1210/en.2007-0271

Takeshima, H., Wakabayashi, M., Hattori, N., Yamashita, S., Ushijima, T., 2015. Identification of coexistence of DNA methylation and H3K27me3 specifically in cancer cells as a promising target for epigenetic therapy. Carcinogenesis 36, 192–201. https://doi.org/10.1093/carcin/bgu238

Tanasijevic, B., Dai, B., Ezashi, T., Livingston, K., Roberts, R.M., Rasmussen, T.P., 2009. Progressive accumulation of epigenetic heterogeneity during human ES cell culture. Epigenetics 4, 330–338. https://doi.org/10.4161/epi.4.5.9275

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., Baralle, D., Dabir, T., Gener, B., Goudie, D., Homfray, T., Kumar, A., Pilz, D.T., Selicorni, A., Temple, I.K., Van Maldergem, L., Yachelevich, N., Childhood Overgrowth Consortium, van Montfort, R., Rahman, N., 2014. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nat Genet 46, 385–388. https://doi.org/10.1038/ng.2917

Tatton-Brown, K., Zachariou, A., Loveday, C., Renwick, A., Mahamdallie, S., Aksglaede, L., Baralle, D., Barge-Schaapveld, D., Blyth, M., Bouma, M., Breckpot, J., Crabb, B., Dabir, T., Cormier-Daire, V., Fauth, C., Fisher, R., Gener, B., Goudie, D., Homfray, T., Hunter, M., Jorgensen, A., Kant, S.G., Kirally-Borri, C., Koolen, D., Kumar, A., Labilloy, A., Lees, M., Marcelis, C., Mercer, C., Mignot, C., Miller, K., Neas, K., Newbury-Ecob, R., Pilz, D.T., Posmyk, R., Prada, C., Ramsey, K., Randolph, L.M., Selicorni, A., Shears, D., Suri, M., Temple, I.K., Turnpenny, P., Val Maldergem, L., Varghese, V., Veenstra-Knol, H.E., Yachelevich, N., Yates, L., Clinical Assessment of the Utility of Sequencing and Evaluation as a Service (CAUSES) Research Study, Deciphering Developmental Disorders (DDD) Study, Rahman, N., 2018. The Tatton-Brown-Rahman Syndrome: A clinical study of 55 individuals with de novo constitutive DNMT3A variants. Wellcome Open Res 3, 46. https://doi.org/10.12688/wellcomeopenres.14430.1

Tomizawa, S., Kobayashi, H., Watanabe, T., Andrews, S., Hata, K., Kelsey, G., Sasaki, H., 2011. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development 138, 811–820. https://doi.org/10.1242/dev.061416

Tucci, V., Isles, A.R., Kelsey, G., Ferguson-Smith, A.C., Tucci, V., Bartolomei, M.S., Benvenisty, N., Bourc'his, D., Charalambous, M., Dulac, C., Feil, R., Glaser, J., Huelsmann, L., John, R.M., McNamara, G.I., Moorwood, K., Muscatelli, F., Sasaki, H.,

Strassmann, B.I., Vincenz, C., Wilkins, J., Isles, A.R., Kelsey, G., Ferguson-Smith, A.C., 2019. Genomic Imprinting and Physiological Processes in Mammals. Cell 176, 952–965. https://doi.org/10.1016/j.cell.2019.01.043

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., Absher, D.M., Wold, B.J., Myers, R.M., 2013. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 23, 555–567. https://doi.org/10.1101/gr.147942.112

Veland, N., Lu, Y., Hardikar, S., Gaddis, S., Zeng, Y., Liu, B., Estecio, M.R., Takata, Y., Lin, K., Tomida, M.W., Shen, J., Saha, D., Gowher, H., Zhao, H., Chen, T., 2019. DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. Nucleic Acids Res 47, 152–167. https://doi.org/10.1093/nar/gky947

Veselovska, L., Smallwood, S.A., Saadeh, H., Stewart, K.R., Krueger, F., Maupetit-Méhouas, S., Arnaud, P., Tomizawa, S., Andrews, S., Kelsey, G., 2015. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biology 16, 209. https://doi.org/10.1186/s13059-015-0769-z

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Eynde, A.V., Bernard, D., Vanderwinden, J.-M., Bollen, M., Esteller, M., Croce, L.D., Launoit, Y. de, Fuks, F., 2006. The Polycomb group protein EZH2 directly controls DNA methylation. Nature 439, 871–874. https://doi.org/10.1038/nature04431

Walsh, C.P., Chaillet, J.R., Bestor, T.H., 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet 20, 116–117. https://doi.org/10.1038/2413

Wang, X., Paucek, R.D., Gooding, A.R., Brown, Z.Z., Ge, E.J., Muir, T.W., Cech, T.R., 2017. Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. Nature Structural & Molecular Biology 24, 1028–1038. https://doi.org/10.1038/nsmb.3487

Watson, J.D., Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171, 737–738. https://doi.org/10.1038/171737a0

Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., Lemiesz, A.E., Marchione, D.M., Marunde, M.R., Meiners, M.J., Cheek, M.A., Keogh, M.-C., Bareke, E., Djedid, A., Harutyunyan, A.S., Jabado, N., Garcia, B.A., Li, H., Allis, C.D., Majewski, J., Lu, C., 2019. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature 573, 281–286. https://doi.org/10.1038/s41586-019-1534-3

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., Laird, P.W., 2007. Epigenetic stem cell signature in cancer. Nat Genet 39, 157–158. https://doi.org/10.1038/ng1941

Wienholz, B.L., Kareta, M.S., Moarefi, A.H., Gordon, C.A., Ginno, P.A., Chédin, F., 2010. DNMT3L Modulates Significant and Distinct Flanking Sequence Preference for

DNA Methylation by DNMT3A and DNMT3B In Vivo. PLOS Genetics 6, e1001106. https://doi.org/10.1371/journal.pgen.1001106

Wiśniewski, J.R., Zougman, A., Nagaraj, N., Mann, M., 2009. Universal sample preparation method for proteome analysis. Nat Methods 6, 359–362. https://doi.org/10.1038/nmeth.1322

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., Walter, J., 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2, 1–8. https://doi.org/10.1038/ncomms1240

Wu, H., Zeng, H., Lam, R., Tempel, W., Amaya, M.F., Xu, C., Dombrovski, L., Qiu, W., Wang, Y., Min, J., 2011. Structural and Histone Binding Ability Characterizations of Human PWWP Domains. PLOS ONE 6, e18919. https://doi.org/10.1371/journal.pone.0018919

Xavier, M.J., Roman, S.D., Aitken, R.J., Nixon, B., 2019. Transgenerational inheritance: how impacts to the epigenetic and genetic information of parents affect offspring health. Hum Reprod Update 25, 518–540. https://doi.org/10.1093/humupd/dmz017

Xi, S., Geiman, T.M., Briones, V., Guang Tao, Y., Xu, H., Muegge, K., 2009. Lsh participates in DNA methylation and silencing of stem cell genes. Stem Cells 27, 2691–2702. https://doi.org/10.1002/stem.183

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., Ren, B., 2012. Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. Cell 148, 816–831. https://doi.org/10.1016/j.cell.2011.12.035

Xu, G.-L., Bestor, T.H., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., Viegas-Péquignot, E., 1999. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402, 187–191. https://doi.org/10.1038/46052

Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'Amour, J., Bogutz, A.B., Zhang, Y., Zhang, B., Yu, G., Xia, W., Du, Z., Huang, C., Ma, J., Zheng, H., Li, Y., Liu, C., Walker, C.L., Jonasch, E., Lefebvre, L., Wu, M., Lorincz, M.C., Li, W., Li, L., Xie, W., 2019a. SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. Nat Genet 51, 844–856. https://doi.org/10.1038/s41588-019-0398-7

Xu, Q., Xie, W., 2018. Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. Trends in Cell Biology 28, 237–253. https://doi.org/10.1016/j.tcb.2017.10.008

Xu, W., Li, J., Rong, B., Zhao, B., Wang, M., Dai, R., Chen, Q., Liu, H., Gu, Z., Liu, S., Guo, R., Shen, H., Wu, F., Lan, F., 2020. DNMT3A reads and connects histone H3K36me2 to DNA methylation. Protein Cell 11, 150–154. https://doi.org/10.1007/s13238-019-00672-y

Yagi, M., Kabata, M., Tanaka, A., Ukai, T., Ohta, S., Nakabayashi, K., Shimizu, M., Hata, K., Meissner, A., Yamamoto, T., Yamada, Y., 2020. Identification of distinct loci

for de novo DNA methylation by DNMT3A and DNMT3B during mammalian development. Nat Commun 11, 1–14. https://doi.org/10.1038/s41467-020-16989-w

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., Nitta, K.R., Taipale, M., Popov, A., Ginno, P.A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., Taipale, J., 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356, eaaj2239. https://doi.org/10.1126/science.aaj2239

Yoh, S.M., Lucas, J.S., Jones, K.A., 2008. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev. 22, 3422–3434. https://doi.org/10.1101/gad.1720008

Yokomine, T., Hata, K., Tsudzuki, M., Sasaki, H., 2006. Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. CGR 113, 75–80. https://doi.org/10.1159/000090817

Zhang, Y., Charlton, J., Karnik, R., Beerman, I., Smith, Z.D., Gu, H., Boyle, P., Mi, X., Clement, K., Pop, R., Gnirke, A., Rossi, D.J., Meissner, A., 2018. Targets and genomic constraints of ectopic Dnmt3b expression. eLife 7, e40757. https://doi.org/10.7554/eLife.40757

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., Jeltsch, A., 2010. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res 38, 4246–4253. https://doi.org/10.1093/nar/gkq147

Zhang, Y., Ng, H.-H., Erdjument-Bromage, H., Tempst, P., Bird, A., Reinberg, D., 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 13, 1924–1935.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., Lee, J.T., 2008. Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science 322, 750–756. https://doi.org/10.1126/science.1163045

Ziller, M.J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., Gnirke, A., Meissner, A., 2011. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet 7, e1002389. https://doi.org/10.1371/journal.pgen.1002389

Zwaka, T.P., Thomson, J.A., 2005. A germ cell origin of embryonic stem cells? Development 132, 227–233. https://doi.org/10.1242/dev.01586

10 Appendices

Appendix 1. Single-cell oocyte samples.

Table with sample numbers used in Figure 3.11, individual oocyte sample IDs, group to which they were assigned for pseudo-bulk analysis in Chapter 3.4, number of unique reads in each sample (in million) and percentage of all CpG positions in the genome covered.

Sample number	**Sample ID**	**Assigned pseudo-bulk group**	**No. of reads (mln)**	**CpG coverage (%)**
5	2_+/+_310_1c_1	*+/+_1*	2	5.21
16	3_+/+__310_1d_2	*+/+_1*	1.6	5.07
17	3_+/+__310_1d_3	*+/+_1*	2.1	6.12
32	6_+/+__310_1c_5	*+/+_1*	1	2.41
33	6_+/+__310_1c_6	*+/+_1*	4.4	10.86
15	3_+/+__310_1d_1	*+/+_2*	3.3	9.73
29	6_+/+__310_1c_2	*+/+_2*	1.6	5.10
35	6_+/+__310_1c_8	*+/+_2*	1.4	3.82
38	7_+/+__310_1d_2	*+/+_2*	2.6	7.66
39	7_+/+__310_1d_3	*+/+_2*	1.7	5.32
7	2_+/+_310_1c_7	*+/+_3*	1.8	4.45
21	3_+/+__310_1d_7	*+/+_3*	2.9	8.45
28	6_+/+__310_1c_1	*+/+_3*	1.6	4.53
31	6_+/+__310_1c_4	*+/+_3*	1.2	3.02
34	6_+/+__310_1c_7	*+/+_3*	1.3	3.19

37	7_+/+__310_1d_1	*+/+_3*	2.3	6.44
6	2_+/+_310_1c_5	*+/+_4*	2.1	5.48
18	3_+/+__310_1d_4	*+/+_4*	2.2	6.30
19	3_+/+__310_1d_5	*+/+_4*	1.8	5.15
20	3_+/+__310_1d_6	*+/+_4*	2.6	6.49
30	6_+/+__310_1c_3	*+/+_4*	2	5.10
4	2_+/D329A_310_1b_7	*+/D329A_1*	4	10.07
12	3_+/D329A__310_1e_5	*+/D329A_1*	3.1	8.43
14	3_+/D329A__310_1e_8	*+/D329A_1*	3.1	8.95
22	6_+/D329A__310_1b_1	*+/D329A_1*	1.4	3.85
3	2_+/D329A_310_1b_6	*+/D329A_2*	3.4	8.30
8	3_+/D329A__310_1e_1	*+/D329A_2*	2.6	6.96
10	3_+/D329A__310_1e_3	*+/D329A_2*	3	8.30
36	7_+/D329A__310_1e_2	*+/D329A_2*	2.6	6.98
2	2_+/D329A_310_1b_4	*+/D329A_3*	2.8	7.50
9	3_+/D329A__310_1e_2	*+/D329A_3*	3.5	8.80
24	6_+/D329A__310_1b_3	*+/D329A_3*	1.8	4.89
25	6_+/D329A__310_1b_4	*+/D329A_3*	1.7	4.98
26	6_+/D329A__310_1b_5	*+/D329A_3*	1.8	5.28
1	2_+/D329A_310_1b_3	*+/D329A_4*	3	7.35
11	3_+/D329A__310_1e_4	*+/D329A_4*	1.8	4.63

13	3_+/D329A__310_1e_6	*+/D329A_4*	3.4	9.22
23	6_+/D329A__310_1b_2	*+/D329A_4*	1.6	3.98
27	6_+/D329A__310_1b_6	*+/D329A_4*	2	5.31

Appendix 2. ChIP-seq library summary.

Sample ID	***Genotype***	**ChIP antibody**	**Uniquelly mapped reads**
1_127.1a	+/+	K4me3	26,065,430
2_127.1b	*Δ/D329A*	K4me3	25,531,851
3_127.1c	+/+	K4me3	24,742,605
4_136.7a	+/+	K4me3	25,346,843
5_136.7c	*Δ/D329A*	K4me3	20,052,247
6_136.8a	*Δ/D329A*	K4me3	28,877,523
7_127.1a	+/+	K27me3	23,941,847
8_127.1b	*Δ/D329A*	K27me3	23,927,549
9_127.1c	+/+	K27me3	25,697,305
10_136.7a	+/+	K27me3	30,991,066
11_136.7c	*Δ/D329A*	K27me3	39,804,568
12_136.8a	*Δ/D329A*	K27me3	30,309,992
13_127.1a	+/+	K36me3	32,880,597
14_127.1b	*Δ/D329A*	K36me3	29,937,232
15_127.1c	+/+	K36me3	29,471,102
16_136.7a	+/+	K36me3	28,734,886
17_136.7c	*Δ/D329A*	K36me3	30,238,060
18_136.8a	*Δ/D329A*	K36me3	28,528,176
20_127.1a	+/+	input	18,726,261

21_127.1b	*Δ/D329A*	input	39,855,286
22_127.1c	+/+	input	28,176,825
23_136.7a	+/+	input	30,719,847
24_136.7c	*Δ/D329A*	input	32,097,044
25_136.8a	*Δ/D329A*	input	28,336,929

Appendix 3. bioDNMT3A-interacting proteins in mouse ESCs.

Table of top 20 proteins showing significant interaction with bioDNMT3A1 bait, in comparison to wild-type ESCs.

Protein ID	**Log_2 Fold change**	**adj. p-value**	**Mean intensity Wild-type (n)**	**Mean intensity bioD3A1 (n)**	**Genes**	**Protein**
Q80UW8 RPAB1	-4.68	2.15E-04	21.57 (3)	26.25 (3)	*Polr2e*	DNA-directed RNA polymerases I, II, and III subunit RPABC1 (RNA polymerases I, II, and III subunit ABC1) (DNA-directed RNA polymerase II subunit E) (RPB5 homolog)
P62274 RS29	-4.18	1.11E-02	20.80	24.98	*Rps29*	40S ribosomal protein S29
P62876 RPAB5	-4.14	4.72E-03	20.86 (1)	25.00 (3)	*Polr2l*	DNA-directed RNA polymerases I, II, and III subunit RPABC5 (RNA polymerases I, II,

						and III subunit ABC5) (DNA-directed RNA polymerase III subunit L) (RPB10 homolog)
Q9D0T1 NH2L1	-4.09	9.61E-04	24.60 (4)	28.69 (3)	*Snu13* *Nhp2l1* *Ssfa1*	NHP2-like protein 1 (Fertilization antigen 1) (FA-1) (High mobility group-like nuclear protein 2 homolog 1) (Sperm-specific antigen 1) (U4/U6.U5 small nuclear ribonucleoprotein SNU13) (U4/U6.U5 tri-snRNP 15.5 kDa protein) [Cleaved into: NHP2-like protein 1, N-terminally processed]
P63168 DYL1	-3.99	1.16E-03	23.03 (4)	27.02 (3)	*Dynll1* *Dlc1* *Dncl1* *Dnclc1*	Dynein light chain 1, cytoplasmic (8 kDa dynein light chain) (DLC8) (Dynein light chain LC8-type 1) (Protein inhibitor of neuronal nitric

						oxide synthase) (PIN) (mPIN)
P70372 ELAV1	-3.99	2.15E-04	21.84 (4)	25.83 (3)	*Elavl1* *Elra* *Hua*	ELAV-like protein 1 (Elav-like generic protein) (Hu-antigen R) (HuR) (MelG)
P60867 RS20	-3.97	1.43E-03	21.96 (3)	25.93 (3)	*Rps20*	40S ribosomal protein S20
Q9Z2N8 ACL6A	-3.84	2.75E-03	21.37 (3)	25.21 (3)	*Actl6a* *Actl6* *Baf53a*	Actin-like protein 6A (53 kDa BRG1-associated factor A) (Actin-related protein Baf53a) (BRG1-associated factor 53A) (BAF53A)
Q9Z2D8 MBD3	-3.80	1.63E-03	21.36 (2)	25.17 (3)	*Mbd3*	Methyl-CpG-binding domain protein 3 (Methyl-CpG-binding protein MBD3)
Q9CWR8 DNM3L	-3.70	2.08E-03	27.25 (4)	30.96 (3)	*Dnmt3l*	DNA (cytosine-5)-methyltransferase 3-like

Q9CQS2 NOP10	-3.68	1.79E-03	23.83 (4)	27.51 (3)	*Nop10* *Nola3*	H/ACA ribonucleoprotein complex subunit 3 (Nucleolar protein 10) (Nucleolar protein family A member 3) (snoRNP protein NOP10)
O88508 DNM3A	-3.63	6.74E-03	26.03 (4)	29.65 (3)	*Dnmt3a*	DNA (cytosine-5)-methyltransferase 3A (Dnmt3a) (EC 2.1.1.37) (DNA methyltransferase MmuIIIA) (DNA MTase MmuIIIA) (M.MmuIIIA)
P59708 SF3B6	-3.59	2.94E-03	21.97 (4)	25.55 (3)	*Sf3b6* *Sf3b14*	Splicing factor 3B subunit 6 (Pre-mRNA branch site protein p14) (SF3b 14 kDa subunit)
Q8CCG4 DPPA4	-3.51	5.07E-04	23.69 (4)	27.20 (3)	*Dppa4*	Developmental pluripotency-associated protein 4 (Embryonic stem cell-associated transcript 15-1

						protein) (ECAT15-1)
P62082 RS7	-3.50	8.58E-03	23.22 (2)	26.72 (3)	*Rps7*	40S ribosomal protein S7
P62264 RS14	-3.43	2.08E-03	23.60 (4)	27.03 (3)	*Rps14*	40S ribosomal protein S14
Q64012 RALY	-3.37	6.36E-04	23.93 (4)	27.30 (3)	*Raly Merc*	RNA-binding protein Raly (Maternally-expressed hnRNP C-related protein) (hnRNP associated with lethal yellow protein)
Q3THW5 H2AV	-3.35	8.26E-04	27.84 (4)	31.20 (3)	*H2az2 H2afv H2av*	Histone H2A.V (H2A.F/Z) (H2A.Z variant histone 2)
Q9D7M8 RPB4	-3.34	1.63E-03	22.91 (4)	26.25 (3)	*Polr2d*	DNA-directed RNA polymerase II subunit RPB4 (RNA polymerase II subunit B4) (DNA-directed RNA polymerase II subunit D)

Q9DC33 HM20A	-3.28	1.16E-03	22.26 (4)	25.53 (3)	*Hmg20a* *Ibraf*	High mobility group protein 20A (HMG box-containing protein 20A) (HMG domain-containing protein HMGX1) (Inhibitor of BRAF35) (iBRAF)